

РАЗРАБОТКА АЛГОРИТМА ЯЗЫКОВОЙ ДИАРИЗАЦИИ ДЛЯ РЕШЕНИЯ ПРОБЛЕМЫ МУЛЬТИЯЗЫЧНОСТИ В РАСПОЗНАВАНИИ РЕЧИ

Марчевский В.Д. (ИТМО)

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

Введение. Речевые технологии представляют собой стремительно развивающуюся сферу исследований, открывающую широкие перспективы для создания человеко-машинных интерфейсов. Одним из вызовов в разработке систем автоматического распознавания речи является переключение языков. В данной работе предлагается подход к решению этой проблемы через применение модуля языковой диаризации.

Основная часть. Проблема мультязычности в системах автоматического распознавания речи (ASR) представляет собой многогранный вызов, затрагивающий как аспекты обучения универсальных моделей, так и применение специализированных алгоритмов для языковой детекции и диаризации [1]. Основным подходом является обучение так называемого “переключаемого” ASR, то есть модели, которая способна обрабатывать сразу несколько языков [2]. Диаризация разговорного языка - это задача автоматического разделения и маркировки одноязычных сегментов, присутствующих в речи с переключением языков. Попытки исследования этой задачи в литературе ограничены. Однако, существует ряд работ, которые основаны на обнаружении переключения языка [3] и на идентификации языка вложенных произношений [4]. Алгоритм мультязычного распознавания речи на основе использования модуля языковой диаризации состоит из двух этапов: определение моноязычных сегментов речи и применение соответствующего моноязычного алгоритма ASR. В данной работе разработан алгоритм языковой диаризации, который основывается на методе спектральной кластеризации языковых вложений, вычисленных в скользящем окне, а также на подходе НВР для расчета матрицы сходства. Такой алгоритм выполняет прогнозирование языковых меток для ограниченного числа целевых языков. Ошибка диаризации на тестовом множестве базы данных DISPLACE, содержащей английский язык и хинди с переключением для данного алгоритма составила 27,3%.

Выводы. Представленный подход к языковой диаризации позволяет эффективно решать проблему мультязычности в автоматическом распознавании речи. Разработанный алгоритм показал хорошие результаты на тестовом наборе данных конкурса DISPLACE 2024 (ошибка диаризации составила 27.3%).

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. Lyu D. C., Chng E. S., Li H. Language diarization for conversational code-switch speech with pronunciation dictionary adaptation // 2013 IEEE China Summit and International Conference on Signal and Information Processing. – IEEE, 2013. – С. 147-150.
4. Shah S., Sitaram S., Mehta R. First workshop on speech processing for code-switching in multilingual communities: Shared task on code-switched spoken language identification // WSTCSMC, 2020. – С. 24.