

ИСПОЛЬЗОВАНИЕ ФОНЕТИЧЕСКИХ ПРИЗНАКОВ В ЗАДАЧЕ ГЕНЕРАЦИИ АВТОМАТИЧЕСКОЙ ТРАНСКРИПЦИИ

Ананьева А. Д.¹

Научный руководитель – Томилов А. А.²

¹Университет ИТМО, ²ООО «ЦРТ»

464995@niuitmo.ru

Введение

Генерация автоматической транскрипции остается одной из актуальных задач речевой обработки. Существует ряд сложностей при работе с транскрипцией, одной из которых является высокая вариативность фонетических систем и различия в наборах используемых символов. Это затрудняет перенос моделей между языками и сопоставление результатов различных исследований. Описание фонемы через набор артикуляционных признаков позволяет представить систему языка в универсальном пространстве.

Основная часть

Исследование проводилось на данных корпуса CORPRES, содержащего речевые записи с фонетической разметкой. В работе была использована модель HubertSoft [1], для вычисления эмбедингов по каждому 20мс входного сигнала. Для каждого сегмента входного сигнала определяется соответствующий фонетический символ, что позволило сформировать обучающую выборку вида «акустический вектор - фонетическая единица»

Преобразование фонетической разметки в разметку по признакам осуществлялось с помощью библиотеки PanPhon [2], которая реализует отображение символов IPA в фиксированный вектор характеристик. В исходном представлении каждый признак принимает одно из трёх значений: признак выражен, признак отсутствует и признак.

В качестве базовой модели используется линейный классификатор поверх эмбедингов HubertSoft. Модель обучается предсказывать значения признаков независимо для каждого временного шага. Такое решение позволяет оценить, в какой степени артикуляционная информация уже присутствует в предобученных акустических представлениях.

Качество классификации оценивается с использованием метрики Phonetic Feature Error Rate (PFER) [3], отражающей долю неверно предсказанных признаков. В отличие от фонемной точности, эта метрика позволяет выявить, какие именно артикуляционные параметры представляют наибольшую сложность для модели. Это создаёт возможность более детального анализа структуры акустических эмбедингов и их соответствия описанию.

Выводы

В работе реализован подход к автоматической транскрипции, основанный на предсказании артикуляционных признаков вместо фонемных классов. Переход к признаковому представлению обеспечивает более интерпретируемую оценку качества и позволяет детально анализировать ошибки модели по отдельным артикуляционным параметрам с использованием метрики PFER.

Литература

1. Van Niekerk B. et al. A comparison of discrete and soft speech units for improved voice conversion // ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2022. – С. 6562-6566.

2. Mortensen D. R. et al. Panphon: A resource for mapping IPA segments to articulatory feature vectors //Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers. – 2016. – C. 3475-3484.
3. Taguchi C. et al. Universal automatic phonetic transcription into the international phonetic alphabet //arXiv preprint arXiv:2308.03917. – 2023.