

Оценка психического состояния говорящего по параметрам устной речи

Багрянов Владислав Андреевич

Московский физико-технический институт

Научный руководитель: Чучупал В.Я, к.ф.-м.н., ВЦ ФИЦ ИУ РАН

- Проведен анализ существующих подходов
- Обучена базовая модель основанная на анализе спектограмм с помощью сверток и GRU
- Обучена модель на основе Wav2Vec

Текущие результаты

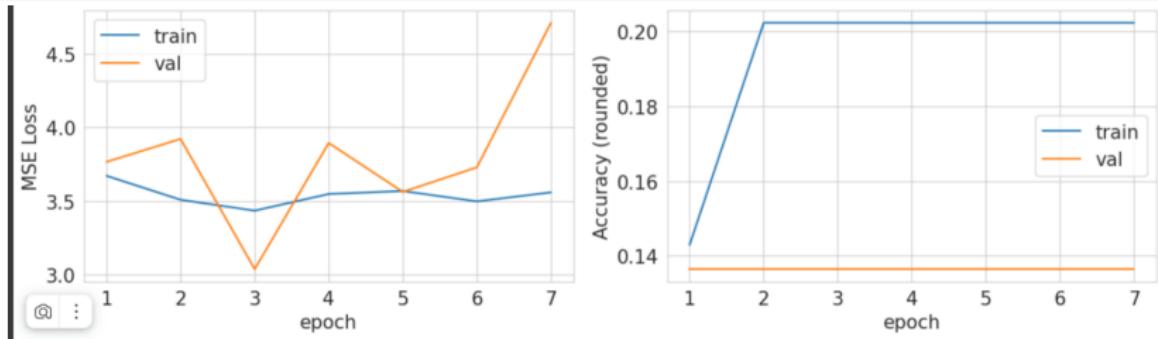


Рис.: Wav2Vec

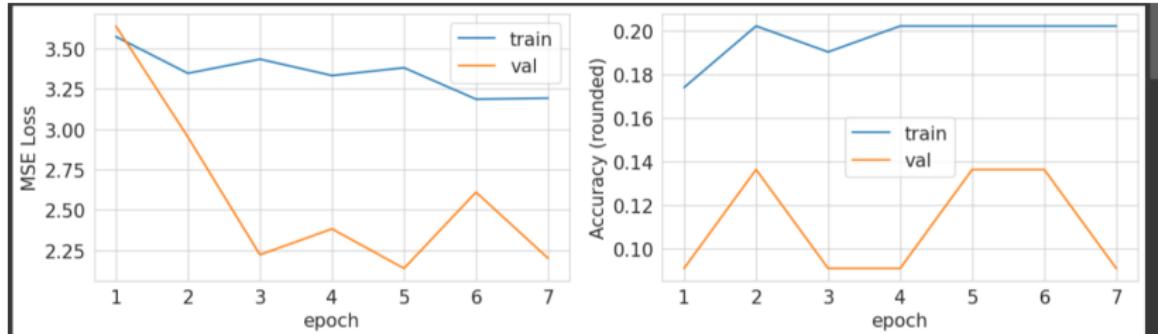


Рис.: Сверточная сеть

Сверточная сеть

Входные данные: Спектограмма

- Свертки с размерностью 32
- Свертка с размерностью 64
- 2 GRU
- Полносвязный слой
- Сигмоида
- Функция активации между слоями - ReLU

Архитектура на основе Wav2Vec2

Входные данные: Wav

- wav2vec2
- 3 Свертки по времени
- Слой Self-attention
- Полносвязный слой
- Сигмоида
- Функция активации между слоями - ReLU

Проблемы:

- На данный момент имеются только данные о здоровых пациентах, из-за чего невозможно провести численные эксперименты
- Модель на основе Wav2Vec оказывается значительно более требовательной к ресурсам

- Получение оставшихся данных и проведение экспериментов
- Обучение и тестирование моделей на основе GigaAM и Whisper
- Оптимизация архитектуры моделей
- Написание статьи