

Оценка психического состояния говорящего по параметрам устной речи

Филиппенков Игорь Андреевич

Московский физико-технический институт

Соавтор: Багрянов В.А

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

- Депрессия является одним из наиболее распространенных психических расстройств
- Традиционные методы диагностики требуют значительных временных и человеческих ресурсов
- Машинное обучение предлагает инструменты для автоматического анализа речи, что позволяет минимизировать субъективность и повысить точность диагностики.

- **Цель исследования:** Создание нейросетевой регрессионной модели, способного предсказывать баллы по шкале Гамильтона на основе речевых данных с использованием модели Wav2Vec 2.0
- **Актуальность:** Результат применения шкал профессиональным психологом все равно весьма субъективный, поэтому использование методов машинного обучения может дать более дешевое решение такого же уровня качества

Общие проблемы данной задачи

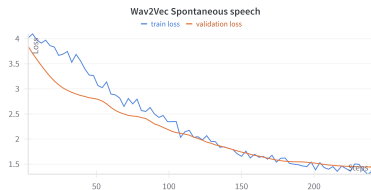
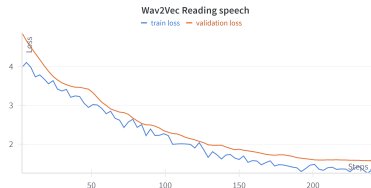
- Разметка медицинских данных требуют профессиональных навыков, поэтому во всех подобных задачах исследователи сталкиваются с большой нехваткой данных
- Оценка уровня депрессии в первую очередь должна опираться на паралингвистических признаках речи, а не текстовой транскрипции.

Существующие подходы и их недостатки

- Методы на основе усредненных признаков речи (eGeMAPs, GMM и др.)
- Использование сверточных нейронных сетей (CNN) и рекуррентных нейронных сетей (RNN)

Постановка задачи обучения

- Используем собственный датасет состоящий из примеров спонтанной речи(описание картинки) и чтения теста, и их оценок по шкале Гамильтона
Выделяются классы депрессии в зависимости от набранных баллов:
 - Норма
 - Лёгкое депрессивное расстройство
 - Депрессивное расстройство средней степени тяжести
 - Депрессивное расстройство тяжелой степени
 - Депрессивное расстройство крайне тяжёлой степени
- На данный момент собраны данные только по первым двум классам
- Используем функцию потерь: MSE
- Метрики качества: accuracy, precision, recall

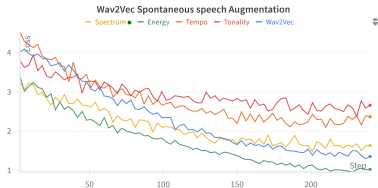
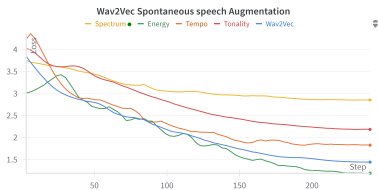


Был проведен сравнительный анализ спонтанной и читаемой речи

Тип речи	Accurasy
Читаемая	0.81
Спонтанная	0.85

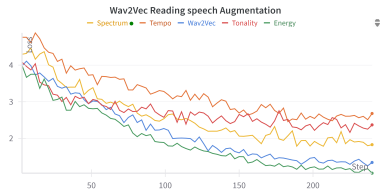
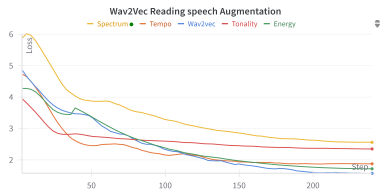
Тип признака	Тип аугментации
Wav2Vec	Без аугментации
Тональные	Pitch shift, Time stretch
Темповые	Speed change, Silence insertion
Энергетические	Volume change, Dynamic range compression
Спектральные	Frequency masking, Time masking

Аугментация



Была проведена аугментация данных спонтанной речи по разным звуковым характеристикам:

Аугментация



Была проведена аугментация данных читаемой речи по разным звуковым характеристикам:

- Обе модели показали высокие результаты наравне с топовыми существующими подходами
- Wav2Vec показал качество немного выше сверточной модели
- Модели лучше обучаются на спонтанной речи
- По результатам аугментации были выявлены наиболее важные признаки

- Обучить модель на оставшейся части датасета
- Обучить другие модели вместо Wav2Vec, например Whisper
- На основе полученных результатов опубликовать статью

- Lu, X., Shi, D., Liu, Y., Yuan, J. (2021). Speech depression recognition based on attentional residual network. *Frontiers in Bioscience (Landmark Edition)*, 26(12), 1746–1759.
- Chen, X., Pan, Z. (2021). A convenient and low-cost model of depression screening and early warning based on voice data using for public mental health. *International Journal of Environmental Research and Public Health*, 18(12), 6441.
- Zhang, L., Duvvuri, R., Chandra, K. K. L., Nguyen, T., Ghomi, R. H. (2020). Automated voice biomarkers for depression symptoms using an online cross-sectional data collection initiative. *Depression and Anxiety*, 37(7), 657–669.
- Donaghy, P., Ennis, E., Mulvenna, M. et al. A Review of Studies Using Machine Learning to Detect Voice Biomarkers for Depression. *J. technol. behav. sci.* (2024).