УДК 004.8

ОБЗОР СОВРЕМЕННЫХ МЕТОДОВ И МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ ДИЗАРТРИЧЕСКОЙ РЕЧИ

Пивоварова А. А.[1]
Научный руководитель – д. т. н. Матвеев Ю. Н.[1]
[1]Университет ИТМО
aapivovarova@itmo.ru

Введение

Дизартрия - это расстройство речи, приводящее к нарушениям артикуляции, фонации и просодики. Автоматическая диагностика дизартрии и оценка степени её тяжести являются актуальными задачами, поскольку традиционные клинические методы субъективны, трудоёмки и требуют привлечения квалифицированных специалистов [1]. Данная работа представляет обзор основных методов машинного обучения, применяемых для бинарной классификации (диагностика наличия дизартрии) и множественной классификации (определение типа дизартрии или степени её тяжести). На основе проведённого анализа предлагается методика бинарной классификации дизартрии с использованием дообучения фундаментальной модели речи (SFM) Wav2Vec2-BERT с параметро-эффективной настройкой (PEFT).

Основная часть

Анализ литературы последних лет показывает эволюцию подходов к автоматической классификации дизартрии.

С развитием глубокого обучения большее распространение получили архитектуры на основе свёрточных нейронных сетей (CNN) и рекуррентных моделей. Такие работы, как [2; 3; 4], провели сравнительные исследования архитектур на основе CNN для детекции и классификации тяжести дизартрии.

В наше время, наиболее перспективным направлением является применение предобученных SFM. В работах [5; 6; 7] авторами была продемонстрирована эффективность дообучения SFM методами PEFT для задач классификации дизартрии.

Несмотря на впечатляющие результаты, в существующих исследованиях выделяется ряд нерешённых проблем.

Разные исследования используют различные протоколы оценки, метрики и критерии включения [1]. Данные часто бывают несбалансированы - к примеру, в работе [5] при формировании подвыборок пациентов по уровням разборчивости речи соблюдается гендерный дисбаланс. В дополнение к этому, авторы таких работ, как [5; 7], тестируют подходы на одном языке и одном датасете, что не позволяет оценить обобщающую способность моделей на разных языках и типах дизартрии.

Также необходимо заметить, что в работах [5; 7] используются либо не оптимизированные для задач классификации SFM (Whisper), либо не самые современные версии доступных архитектур (Wav2vec2-base). Существует более современная архитектура Wav2Vec2-BERT, которая предполагает иную стратегию заморозки слоёв, нежели описанная в литературе. В [6] используется Wav2Vec2-BERT для классификации тяжести дизартрии, однако авторы не раскрывают детали процесса.

Для преодоления указанных ограничений предлагается следующий экспериментальный подход. В качестве основы выбирается модель Wav2Vec2-BERT. Предлагается экспериментировать со следующими стратегиями: обучение только классификационной головы (“linear probing”); заморозка всех слоёв, кроме встроенных адаптеров и классификатора; дообучение модели через метод LoRA.

Для тестирования подхода выбраны три датасета: TORGO (английский язык), EasyCall (итальянский язык), и MSDM (язык мандарин).

Выводы

Анализ литературы показывает эволюцию методов классификации дизартрии от традиционных подходов к глубоким нейросетевым архитектурам и современным SFM. При этом трансформерные архитектуры демонстрируют наилучшие результаты при наличии ограниченных данных. Ключевыми проблемами в вопросе автоматической классификации дизартрической речи остаются малое количество доступных аудиоданных с дизартрической и прочей патологической речью. Наиболее перспективным направлением представляется использование предобученной Wav2vec2-BERT с параметро-эффективным дообучением. Предложенная методика будет протестирована на трёх разнородных датасетах, что позволит оценить её обобщающую способность для различных типов дизартрии, языков и условий записи.

Литература

1. Sindhu, I. Automatic Speech and Voice Disorder Detection Using Deep Learning—A Systematic Literature Review / I. Sindhu, M.S. Sainin – Текст : непосредственный. // IEEE Access. 2024. Т. 12. – С. 49667-49681.

2. Joshy, A. Automated Dysarthria Severity Classification: A Study on Acoustic Features and Deep Learning Techniques / A. Joshy, R. Rajan – Текст : непосредственный. // IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2022. Т. PP. Automated Dysarthria Severity Classification. – С. 1-1.

3. Shih, D.-H. Dysarthria Speech Detection Using Convolutional Neural Networks with Gated Recurrent Unit // Healthcare. 2022. Vol. 10. № 10. – P. 1956.

4. Sajiha, S. Automatic dysarthria detection and severity level assessment using CWT-layered CNN model // EURASIP Journal on Audio, Speech, and Music Processing. 2024. Vol. 2024. № 1. – P. 33.

5. Chowdary, P. N. A Few-Shot Approach to Dysarthric Speech Intelligibility Level Classification Using Transformers // 2023 14th International Conference on Computing Communication and Networking Technologies. – Delhi, India: IEEE, 2023. – P. 1-6.

6. Dai, W. Fine-Tuning Pre-Trained Audio Models for Dysarthria Severity Classification: A Second Place Solution in the Multimodal Dysarthria Severity Classification Challenge // 2024 IEEE 14th International Symposium on Chinese Spoken Language Processing (ISCSLP). – Beijing, China: IEEE, 2024. Fine-Tuning Pre-Trained Audio Models for Dysarthria Severity Classification. – P. 151-153.

7. Purohit, T. Automatic Parkinson’s disease detection from speech: Layer selection vs adaptation of foundation models // ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – Hyderabad, India: IEEE, 2025. Automatic Parkinson’s disease detection from speech. – P. 1-5.