Нейросетевой анализ акустических различий носителей и изучающих русский язык

Черемискин Егор Андреевич

Студент

Факультет ВМК МГУ имени М.В.Ломоносова, Москва, Россия

E-mail: egorcheremiskin@yandex.ru

Научный руководитель *— Гуров Сергей Исаевич*

В работе рассматривается задача нейросетевого анализа русской речи носителей языка и китайскоговорящих изучающих с целью выявления устойчивых акустических различий в произношении. В отличие от большинства исследований, ориентированных либо на автоматическое распознавание речи, либо на оценку произношения, в данной работе предлагается комплексная постановка, объединяющая две задачи: классификацию типа диктора (носитель / изучающий) и распознавание произнесённых слов (audio2text). Такой подход позволяет анализировать не только итоговую точность распознавания, но и структуру ошибок, связанную с фонетическими особенностями языка и влиянием иностранного акцента.

В качестве признаков использовались как традиционные акустические характеристики, включая мел-частотные кепстральные коэффициенты, так и современные нейросетевые векторные представления речи. На основе данных представлений обучались модели машинного обучения и нейронные сети для решения задач классификации дикторов и распознавания слов. Такой подход позволяет учитывать не только спектральные свойства сигнала, но и его временную динамику, контекст и скрытые фонетические закономерности.

Экспериментальные результаты показали, что нейросетевые представления речи существенно превосходят классические спектральные признаки как в задаче классификации, так и в задаче распознавания слов. При использовании мел-частотных кепстральных коэффициентов точность классификации составила 69%, тогда как модели на основе векторных представлений wav2vec 2.0 достигли 82% точности.

Дополнительный прирост качества наблюдается при учёте временной динамики сигнала. Последовательные модели, обрабатывающие полную временную структуру слова, обеспечили увеличение точности до 87%. Наилучшие результаты были получены при использовании механизма внимания, позволяющего выделять наибо-

лее информативные фрагменты речи: итоговая точность классификации достигла 90% при ROC-AUC около 0.94. Это свидетельствует о важности не только спектральных характеристик, но и временной организации произношения при анализе иностранного акцента.

Анализ ошибок автоматического распознавания показал наличие систематических отклонений в произношении у изучающих русский язык, связанных с переносом фонетических особенностей родного языка. Предложенный комплексный нейросетевой подход позволяет не только различать тип диктора, но и выявлять акустические закономерности, характерные для неродной речи, что может быть использовано в задачах лингвистики, автоматической оценки произношения и систем обучения иностранным языкам.

Литература

1. Amodei D. et al. Deep Speech 2: End-to-End Speech Recognition in English and Mandarin // ICML, 2016.
2. Schuller B., Batliner A. Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. Wiley, 2014.
3. Derwing T., Munro M. Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research. John Benjamins, 2015.