VoiceBridge

Severity-Aware Dysarthric Speech Analysis Using Conformer Networks and GAN-Based Data Augmentation

1. Project Overview

VoiceBridge is a research-oriented deep learning framework designed to improve dysarthric speech analysis. The project focuses on accurate dysarthria severity classification and speech intelligibility improvement by combining Conformer-based speech modeling with carefully selected GAN-based data augmentation.

2. Motivation

Dysarthric speech datasets suffer from limited data availability and severe class imbalance, particularly for severe dysarthria. Traditional models often fail to generalize well under these conditions. VoiceBridge addresses this problem by introducing a two-phase framework that fairly evaluates multiple GAN models before applying data augmentation.

3. Dataset

The TORGO dysarthric speech dataset is used in this project. It contains speech samples categorized into mild, moderate, and severe dysarthria levels. A speaker-independent train, validation, and test split is applied to ensure fair evaluation.

4. Phase 0: Exploratory Data Analysis (EDA)

EDA is performed to understand dataset characteristics and justify data augmentation. The analysis includes speaker-wise distribution, severity-wise imbalance, utterance duration statistics, and spectrogram visualization. The findings confirm data scarcity and imbalance, especially for severe dysarthria.

5. Audio Preprocessing and Feature Extraction

All audio signals are resampled to 16 kHz, silence is trimmed, and amplitude normalization is applied. Log-Mel spectrograms are used as the primary feature representation, while MFCCs are used for baseline comparison.

6. Phase 1: Conformer Baseline and GAN Comparison

In Phase 1, a Conformer encoder with a softmax classifier is trained as a baseline severity classification model. The model is evaluated using F1-score for classification and WER, CER, and SER for speech intelligibility.

Three GAN models are then evaluated independently using only GAN-generated dysarthric speech: SpecGAN, Conditional GAN (cGAN), and CycleGAN. Each GAN is assessed using the same metrics to ensure fairness.

7. Phase 2: Best-GAN-Based Data Augmentation

The best-performing GAN from Phase 1 is selected based on classification accuracy, intelligibility metrics, and consistency across severity levels. The selected GAN is then used to augment the original TORGO dataset in a severity-aware manner.

The Conformer model is retrained using the augmented dataset while keeping the architecture and training configuration unchanged to ensure fair comparison.

8. Final Evaluation and Results

Final evaluation is performed using F1-score, WER, CER, and SER. Results are reported both overall and severity-wise. The largest performance gains are observed for severe dysarthria, demonstrating the effectiveness of the proposed approach.

9. Key Contributions

- Two-phase severity-aware dysarthric speech analysis framework
- Conformer-based severity classification
- Comparative and unbiased evaluation of multiple GAN models
- Best-GAN selection strategy
- Joint evaluation using classification and intelligibility metrics
- Significant improvements for severe dysarthria

10. Conclusion

VoiceBridge demonstrates that careful GAN selection combined with severity-aware data augmentation and Conformer-based modeling significantly improves dysarthric speech analysis. The framework is robust, reproducible, and suitable for real-world assistive speech applications.