Proposal of self-supervised learning pipeline for dysarthric speech and continuous learning integration

To design a self-supervised learning pipeline that caters to dysarthric speech, we have to consider the unique characteristics of dysarthric speeches that includes slurred articulation, irregularities in speech rate, and possible phonemic distortions.

Pretraining Model Phase:

In this phase, the pipeline can leverage self-supervised learning (SSL) frameworks. Examples like contrastive and masked region prediction can be used to pretrain the model on a large dataset of regular non dysarthric speeches. LFB2Vec is an SSL technique that can be used to learn meaningful representations without explicit annotations.

Finetuning Model Phase:

After obtaining a pretrained model, it will be further finetuned by integrating techniques such as noise reduction, segmentation, and feature extraction. Additionally, an accurate Audio Event Detection (AED) model can be incorporated to filter background noises and enhance quality of the input audio. We should choose a reliable optimizer such as AdamW to provide stability in learning rate and a loss function of flatNCE adapted from various deep learning domains that proves useful in addressing biased estimations is paramount in improving SSL performance for dysarthric speech usage.

Continuous Learning Integration:

1. Continuous Data Collection

In the integration of continuous learning, data of dysarthric speech can be continuously collected to further refine the model. Platforms such as speech therapy clinics and applications designed for individuals with speech impairment are ideal for continuously collecting data of dysarthric speeches as they provide reliable steam of data and a controlled environment for data processing.

2. Model Evaluation Feedback Loop

A feedback loop mechanism can be incorporated into the model where the model will be further refined and trained on the newly collected dysarthric speech data. The model's performance will be monitored and evaluated dynamically, and the results will be used to gain insights to identify areas of improvement.

3. Model Update Feedback Loop

The model will undergo adaptive changes to integrate new insights based on the results of the previous continuous model evaluation and changes in data distribution from new sources of dysarthric speech. This involves finetuning the hyperparameters of the model, adjusting or updating the model's architecture to accommodate to new dysarthric speech patterns.

In summary, this proposed self-supervised learning pipeline that caters to dysarthric speeches involves pretraining and obtaining a robust ASR using SSL frameworks, followed by finetuning the model through noise reduction and segmentation techniques to accommodate to noisy data and potential dysarthric speech data. Finally, continuous learning mechanisms will be incorporated to adapt the model to evolving dysarthric speech nuances. This adaptive approach aims to improve the model's performance continuously to adopt new variations in dysarthric speeches overtime, further improving the accuracy and overall effectiveness in recognizing and transcribing dysarthric speeches.

References:

https://arxiv.org/pdf/2205.08598.pdf

https://arxiv.org/pdf/2203.15431.pdf

Integrity Disclaimer: AI is used here to summarize article contents to draw insights and generate inspiration.