Exploring Pre-Trained Automatic Speech Recognition Model HuBERT for Early Alzheimer's Disease and Mild Cognitive Impairment Detection in Speech

Authors : Monica Gonzalez Machorro

Abstract : Dementia is hard to diagnose because of the lack of early physical symptoms. Early dementia recognition is key to improving the living condition of patients. Speech technology is considered a valuable biomarker for this challenge. Recent works have utilized conventional acoustic features and machine learning methods to detect dementia in speech. BERT-like classifiers have reported the most promising performance. One constraint, nonetheless, is that these studies are either based on human transcripts or on transcripts produced by automatic speech recognition (ASR) systems. This research contribution is to explore a method that does not require transcriptions to detect early Alzheimer's disease (AD) and mild cognitive impairment (MCI). This is achieved by fine-tuning a pre-trained ASR model for the downstream early AD and MCI tasks. To do so, a subset of the thoroughly studied Pitt Corpus is customized. The subset is balanced for class, age, and gender. Data processing also involves cropping the samples into 10-second segments. For comparison purposes, a baseline model is defined by training and testing a Random Forest with 20 extracted acoustic features using the librosa library implemented in Python. These are: zero-crossing rate, MFCCs, spectral bandwidth, spectral centroid, root mean square, and short-time Fourier transform. The baseline model achieved a 58% accuracy. To fine-tune HuBERT as a classifier, an average pooling strategy is employed to merge the 3D representations from audio into 2D representations, and a linear layer is added. The pre-trained model used is 'hubert-large-ls960-ft'. Empirically, the number of epochs selected is 5, and the batch size defined is 1. Experiments show that our proposed method reaches a 69% balanced accuracy. This suggests that the linguistic and speech information encoded in the self-supervised ASR-based model is able to learn acoustic cues of AD and MCI.

Keywords : automatic speech recognition, early Alzheimer's recognition, mild cognitive impairment, speech impairment

Conference Title : ICPSP 2022 : International Conference on Phonetics and Speech Processing

Conference Location : London, United Kingdom

Conference Dates : March 11-12, 2022