ELEC-E5510 - Project Plan

Hung Nguyen
Joona Sorjonen

Contents

1 The problem

The problem is a speech recognition task using data from the Common Voice Corpus for Esperanto voice samples, which have been additionally obfuscated with random character changes.

2 Data

Training data is 6000 training samples (from multiple speakers), and 1000 test samples.

Data has been additionally obfuscated by making random character changes into the training data labels (95 % of training samples and 100 % of test samples).

3 Methods and experiments

We chose the python library torchaudio for this task, since it's the one our group has the most experience with. So far, we have experimented with loading the training and test data, and as a proof-of-concept extracted Spectrogram, Mel-Spectrogram and MFCCs features from the data.

We experimented with extracting features using a very simple pretrained (not finetuned) CTC model, using Wav2Vec. The model produced overall poor results, however the resulting output does resemble the correct transction, so with actual training this model might be suitable for the task.

For the decoder, we tested with a very naive decoder, which simply picks the highest probablity from the accoustic model.

4 Tasks

Somebody, Deadline 1.12.2024 Research CTC models, especially training them using our data, try to run a simple proof-of-concept of training the model using our own data.

Somebody, Deadline 1.12.2024 Research ways to improve our decoder (language model), choose a suitable approach to implement (and present it)