HW3 speech recognition

Ido Terner

January 2023

1 Results

The WER and CER results on the train and validation set can be seen in Table 1.

	Train		**Validation**	
Decoder	**WER**	**CER**	**WER**	**CER**
CTC decoder with LM beam 1	0.436	0.447	0.434	0.446
CTC decoder with LM beam 50	0.013	0.012	0.014	0.013
CTC decoder with LM beam 500	0.013	0.012	0.014	0.013
CTC decoder without LM beam 1	0.473	0.486	0.473	0.49
CTC decoder without LM beam 50	0.012	0.01	0.012	0.01
CTC decoder without LM beam 500	0.012	0.01	0.012	0.01
Greedy decoder	0.222	0.061	0.231	0.066

Table 1: The WER, CER results on the train and validation sets for each configuration.

After looking at the results, we chose to use the CTC decoder without language model with a beam size of 50 for the test prediction part, which can be seen in the file *"output.txt"*.

2 Acoustic Model Architecture

The used acoustic model is a pre-trained version of **Wav2Vec 2.0**, which is the base architecture, with an extra linear layer module. The model was pre-trained on 960 hours of unlabeled audio from *LibriSpeech* dataset (the combination of "train-clean-100", "train-clean-360", and "train-other-500"). In addition, fine-tuned for ASR on the same audio with the corresponding transcripts.

The model was fine-tuned for 10 epochs on the given dataset. The full hyper-parameters used in the fine-tuning process can be seen in the file *"hyper-parametrs.json"*, or in *"AcousticModel.py"*.

3 Run The Code

To run the code:

1. Put all the python files in the same directory, in addition to the train, test dataset, *"lexicon.txt"*, and *"train transcription.txt"*.
2. Run the file *"CreateLM.py"* in order to create the *.arpa* language model.
3. If you already have a trained model (i.e. a folder called *"checkpoints"* with a saved checkpoint called *"checkpoint.pt"*) skip to step 5.
4. Run the file *"AcousticModel.py"* in order to train and save the fine-tuned acoustic model.
5. Run the file *"decoding.py"* in order to get the WER and CER results for each configuration. In addition, running this file will result in the creation of *"output.txt"*.