Homework 10 – Speech

Arthur J. Redfern
arthur.redfern@utdallas.edu

---

0 Outline

1 Reading
2 Theory
3 Practice

1 Reading

1. Speech
   Motivation: understand common uses of xNNs in speech applications
   https://github.com/arthurredfern/UT-Dallas-CS-6301-CNNs/blob/master/Lectures/xNNs_100_Speech.pdf

   Complete

2. A comparison of sequence-to-sequence models for speech recognition
   Motivation: an alternative presentation of speech to text transduction covering CTC, RNN transducer and attention based models
   https://www.isca-speech.org/archive/Interspeech_2017/pdfs/0233.PDF

   Complete

3. First-pass large vocabulary continuous speech recognition using bi-directional recurrent DNNs
   Motivation: Beam search is a critical component of accurate speech to text transduction (and language to language translation), but was not covered in detail in the slides. To address this, read the following paper that shows how to incorporate an external language model with speech to text transduction networks using beam search
   https://arxiv.org/abs/1408.2873

   Complete

4. An all-neural on-device speech recognizer

Motivation: Embedded devices are an important target for speech to text transduction systems. Read the following blog post and paper on a deployed speech to text transduction system that includes many items we've discussed: RNN transducer models, weight quantization, beam search, … and a number of items we haven't.
https://ai.googleblog.com/2019/03/an-all-neural-on-device-speech.html

Complete

5. Streaming end-to-end speech recognition for mobile devices
   Motivation: corresponding technical article
   https://arxiv.org/abs/1811.06621

   Complete

2 Theory

None

3 Practice

6. Speech to text transduction. Understand the following example and run it in Google Colab:

   - https://www.assemblyai.com/blog/end-to-end-speech-recognition-pytorch
   - https://colab.research.google.com/drive/1IPpwx4rX32rqHKpLz7dc8sOKspUa-YKO

   Complete

7. Text to speech transduction. Understand the following example and run it in Google Colab:

   - https://pytorch.org/hub/nvidia_deeplearningexamples_tacotron2/

   Complete

6. [Optional] Check out some large speech / audio projects in PyTorch:

   - https://pytorch.org/tutorials/beginner/audio_preprocessing_tutorial.html
   - https://pytorch.org/audio/stable/index.html
   - https://github.com/espnet/espnet
   - https://github.com/mravanelli/pytorch-kaldi

- https://speechbrain.github.io

  Complete