Deep Learning in Speech-to-Text Translation

Tsz Kin Lam

Department of Computational Linguistics
lam@cl.uni-heidelberg.de

17th Nov, 2020

Introduction

Lecturer (Not a Professor): Tsz Kin Lam

- 2nd-year PhD student of Prof. Stefan Riezler
- Concentrations:
  1. (Deep) Machine/Speech Translations
  2. Reinforcement Learning
  3. Bayesian Learning
  4. Multi-agent Learning
- I have no linguistics background
- Office Hour: by appointment or try to talk to me when you meet me.
- E-mail: lam@cl.uni-heidelberg.de

Course Info

Proseminar / Hauptseminar

1. Read all papers to be presented and participate to the discussions, in particular,
   - **prepare $\sim$2 questions for each presentation**
2. Present paper(s)
3. Presenter should also write a summary about questions discussed
4. Implementation Project (Later)
   - Group or individual
   - The topic can be different from the papers presented
   - Be sure that you have access to GPU(s)

Course Info cont.

Presentation and discussion:

- $\sim$ 45 minutes plus Q&A
- Q&A (Not an oral exam): Flexible and interactive discussions
- Please send me your slides (ideally 1 week) at least two days before your presentations.
- Papers $\rightarrow$ ▸ Course webpage

Papers - Offline Speech-To-Text Translation

1. Attention-passing models for robust and data-efficient end-to-end speech translation
   - Direct End-to-End, Cascaded (ASR+MT) and Hybrid systems
2. Phone Features Improve Speech Translation
   - How to use phonetics to condense existing speech features?
3. Curriculum Pre-training for End-to-End Speech Translation
   - A better transfer learning method for AST
4. Consistent Transcription and Translation of Speech
   - A scenario where both transcriptions and translations are needed
5. Dual-decoder Transformer for Joint Automatic Speech Recognition and Multilingual Speech Translation
   - A solution to improve point 4?

Papers - Simultaneous Translation

1. STACL: Simultaneous Translation with Implicit Anticipation and Controllable Latency using Prefix-to-Prefix Framework
   - Simultaneous text-to-text translation
2. Re-translation versus Streaming for Simultaneous Translation
   - What is Re-translation?
3. SimulSpeech: End-to-End Simultaneous Speech to Text Translation
   - The meat

Papers - Speech Representations

Trainable speech representations

1. Neural Discrete Representation Learning
   - VQ-VAE - The beginning of the story
2. vq-wav2vec: Self-Supervised Learning of Discrete Speech Representations
   - How can we apply NLP methods to speech?
3. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations
   - How is it different from the above?
4. Learning Robust and Multilingual Speech Representations
   - This one is **Unsupervised**
5. Towards unsupervised speech recognition and synthesis with quantized speech representation learning
   - Unsupervised too

End

Please sign up by next class.
I will give a tutorial about Seq2Seq next week
We start our discussions on 1st Dec 2020