109-2 Research Topics

March 25, 2021

Speech Representation

- Build a better representation for tasks involving speech

- Benchmark speech representation in a variety of tasks

- Holy grail: Redefine the principle for all the speech processing tasks

  - One universal representation model + simple downstream model and simple training -> all tasks solved

- Schedule:

  - Read papers to follow up the recent trend on speech representation learning (2-3 weeks)

  - Hands-on: **s3prl** toolkit, built by me and many members in our lab, collaborated with Amazon, FB, CMU, and JHU.

    - Pre-training, probing, fine-tuning

    - **The most complete collection** for all the speech representations

    - **Various downstream tasks:** ASR, Speaker verification, diarization, STD, slot filling, intent classification...

    - **Contribute** to the project if you get an good pull request idea (Welcome to discuss with me.)

Speech Representation

- **If you find me suitable for you**, please contact Leo with *leo19941227@gmail.com*

Voice Style Transfer

Voice Style Transfer technologies aim at modifying the voice style of utterance(s), including but not limited to: Voice Conversion (VC), Emotional VC, etc.

In the following weeks, you'll be doing:

- Studying some excellent NN-based VC papers as an entry point (3-4 weeks)
- Training and testing few VC models based on open-source projects (1-3 weeks)
- Implementing an one-to-one or any-to-one VC approach (2 weeks up)

Contact me (林義聖 r08922048@ntu.edu.tw) if these interest you.

Non-autoregressive Speech Translation

- Speech-to-text translation (S2T): speech recognition (ASR) + machine translation (MT), but recent focus is on end-to-end systems, i.e. directly translate speech into text.
- Non-autoregressive (NAR): when generating text, we typically do it one token at a time (autoregressive). Can we do it faster? All tokens at the same time?
- Roadmap
  - Get familiar with NAR: study some essential NAR papers, which are mostly applied to MT.
  - Study the current trend on NAR applied to S2T (only a few papers)
  - Implementation is to be discussed. We can:
  - Use open-source toolkit like fairseq
  - Or implement one together
- Contact: 張致強 r09922057@g.ntu.edu.tw

組別分配

https://forms.gle/cNJeE2t48DEhvd2J8

填寫表單分發志願，大助教晚上會寄信公布結果