Speech Recognition Term Project Proposal

Stacked De-noising Auto-encoders for end-to-end speech recognition

Nanshu Wang, Ian Jiun Lai, Wei Wang

October 2017

1. What project would you like to do?
    - For HMM-based speech recognition, stacked denoising auto-encoders were a good way to improve the performance of almost all systems without requiring supervision
    - Would like to know if it still hold for end-to-end trained models

2. Which are the final goals of your project
    - Build De-noising auto-encoders and end-to-end models and train on Grid dataset
    - Compare the effect features extracted from auto-encoder and other speech feature, eg, MFCC in end-to-end models.
    - Compare different model architectures, eg, train auto-encoder combined with end-to-end models or train them separately.
    - Compare De-noising auto-encoder and normal auto-encoder (without de-noising)

3. Whom would you like to collaborate with?
    Nanshu Wang, Ian Jiun Lai, Wei Wang

4. Do you have any data sets, tools, evaluation metrics, etc. in mind? (you can ask us)
    - Datasets: Grid dataset
    - Tools: Tensorflow
    - Evaluation metrics: WER (speech recognition), SNR (speech de-noising)

5. What computational resources do you need?
    We need GPUs to run end-to-end models.

6. Do you need any extra computational resources that we can provide?
    GPU resource: one GTX 1060 (6GB memory)