

Espnet 教學與實作

交大電機碩一 張詔揚

ESPnet是什麼？

- ESPnet是一個開源語音處理工具包，旨在提供end-to-end的語音識別、語音合成(TTS)和聲碼轉換等功能
- ESPnet使用PyTorch作為深度學習引擎，因此需安裝與之版本對應的PyTorch
- 本次作業使用ESPnet的egs2來進行，因此不需要安裝Kaldi作為front end

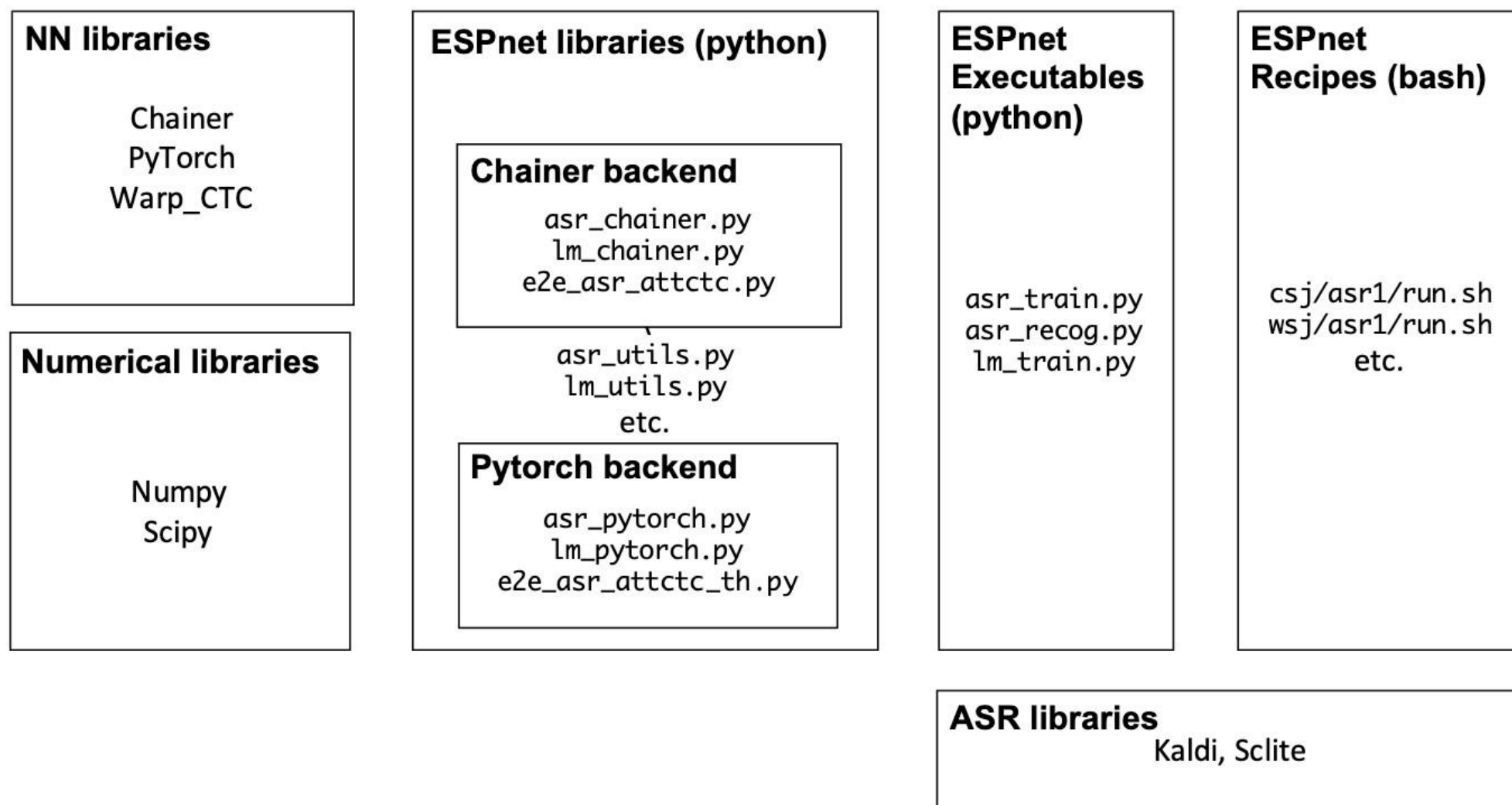


Figure 1: *Software architecture of ESPnet.*

Requirements

- Python 3.6.1+
- gcc 4.9+ for PyTorch1.0.0+
- Cuda 8.0, 9.0, 9.1, 10.0 depending on each DNN library
- Cudnn 6+, 7+
- NCCL 2.0+ (for the use of multi-GPUs)

安裝所需的套件

- `$ sudo apt-get install cmake`
- `$ sudo apt-get install sox`
- `$ sudo apt-get install libsndfile1-dev`
- `$ sudo apt-get install ffmpeg`
- `$ sudo apt-get install flac`

Setup Python environment

- `$ git clone https://github.com/espnet/espnet`
- `$ cd <espnet-root>/tools`
- `$ CONDA_TOOLS_DIR=$(dirname ${CONDA_EXE})/..`
- `$./setup_anaconda.sh ${CONDA_TOOLS_DIR} espnet 3.9`
- `$ make`

訓練資料要放在哪裡？

- espnet/ # Python modules
- egs2/ # The complete recipe for each corpora
- [name]/
- asr1/ # ASR recipe
- downloads/
- resource_aishell/
- -speaker.info
- -lexicon.txt
- data_aishell #放訓練資料
- ...

最重要的一步

- data_aishell/
- wav/
- train/
- -global/ # 訓練音檔
- test/
- -global/ # Kaggles競賽用音檔
- -dev/
- -global/ # 測試音檔(不能是空的)
- transcript/
- aishell_transcript.txt #放與音檔對應的文字檔

開始training

- `$ sudo nvidia-smi -c 3` # GPU獨佔模式
- `$./run.sh --ngpu 2` # 如果你有多張GPU
(開始解bug，可能環境沒設定好，套件沒裝...)
- `./run.sh --stage n` # 從上次中斷的地方執行

查看訓練狀況

- `$tail -20`
`/[username]/espnet/egs2/aishell1/asr1/exp/asr_train_asr_branchfor`
`mer_raw_zh_char_sp/train.log`
- 可以透過`train.log`看看目前的訓練情況
- 或是使用`espnet2.bin package`，測試當前`model`的表現
- 詳見: https://espnet.github.io/espnet/_gen/espnet2.bin.html