

VCC 2020 baseline design

Sian-Yi Chen

Advisor : Tay-Jyi Lin and Chingwei Yeh

Outline

- Action item

- 將 [1] 所提供的 open online resource 建構起來

- Status report

- ◆ 研究

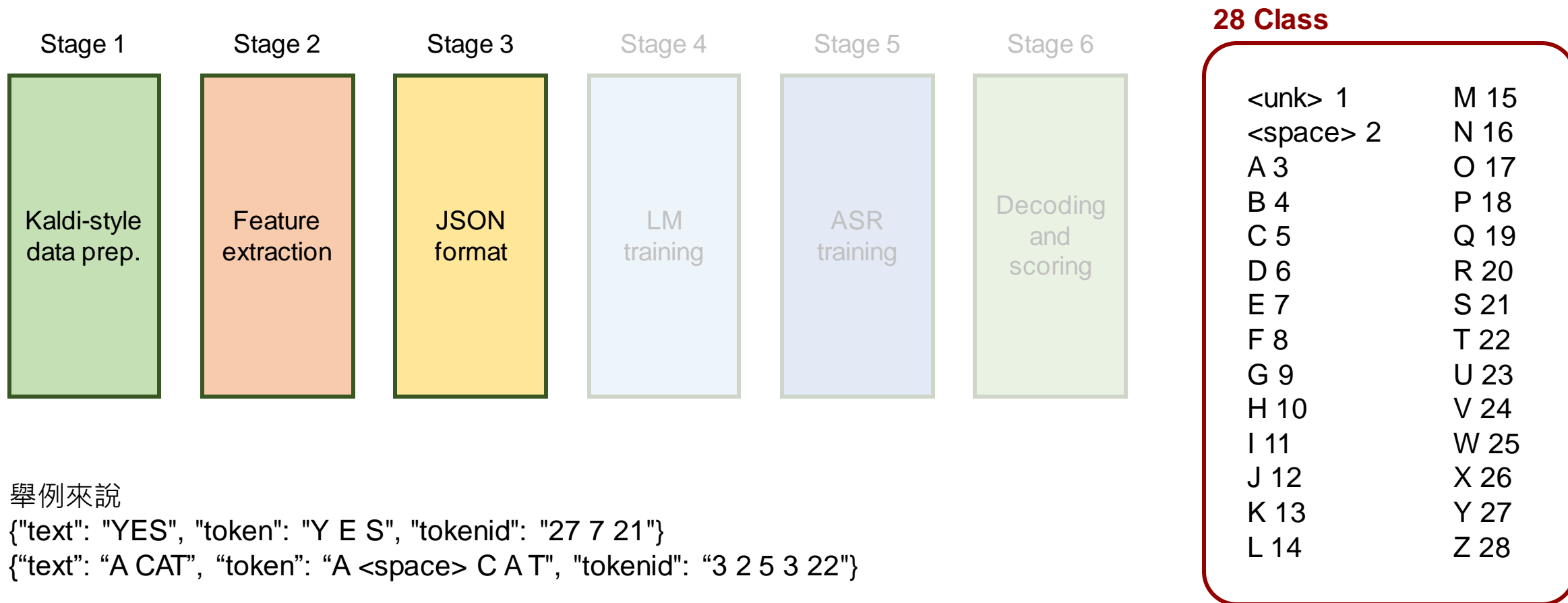
- 目前將需要的環境、套件、語料庫都已經處理好，接著有嘗試使用與論文相同的 LibriSpeech pretrained model，但在執行過程中不確定自己操作的正確性，為了瞭解並熟悉訓練流程有找到 ESPnet 有提供較簡單的 model (an4) 使用，並在後續投影片做介紹。
- 在 ESPnet 提供的各種 model 執行方法大致上皆由 6 個階段完成
 - Stage 1-2：Data preparation (Kaldi-style data & Feature extraction)
 - Stage 3：將語音的 feature 和 transcription 使用 JSON 格式儲存
 - Stage 4-5：訓練神經網路 (Language model & ASR model)
 - Stage 6：解碼以及評估

- ◆ 雜項

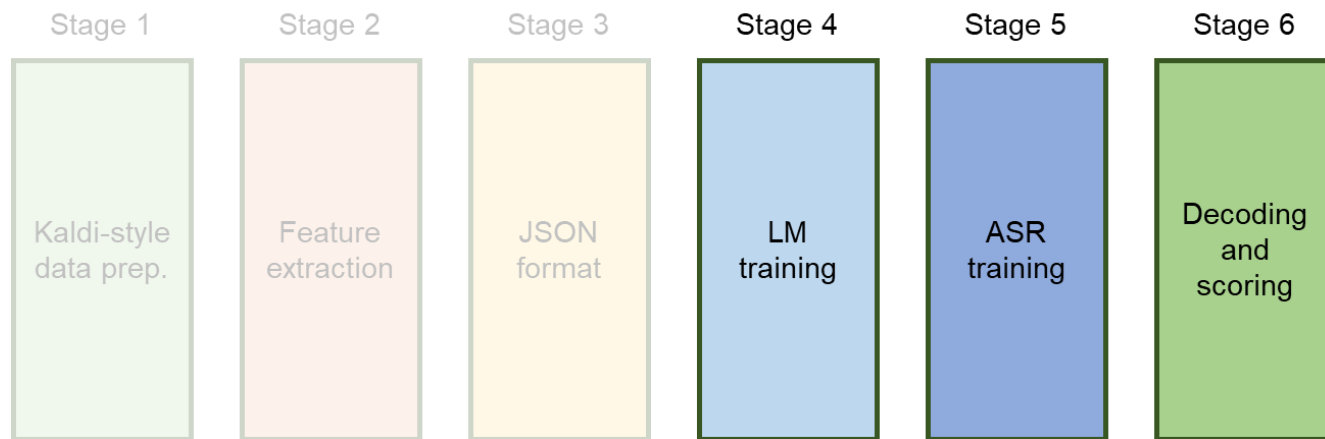
- 上週葉老師對於我說 Kaldi 需要特別多資源覺得很奇怪，我覺得是因為老師說的例子是電機在安裝的時候是在 server 上進行操作，而 server 一般都是 linux 系統，但我們實驗室除了地下室的 powerAI 以外都是 Windows 系統，而在 Windows 系統下安裝 Kaldi 就會變得困難許多。
- 上週 Charlie 學長幫我重灌了 Linux 系統後，我設定好了 SSH 遠端設定，想說如果實驗室有其他想要使用 GPU 資源的可以連線進來使用。

Try run the simple ASR model (an4) provided by ESPnet

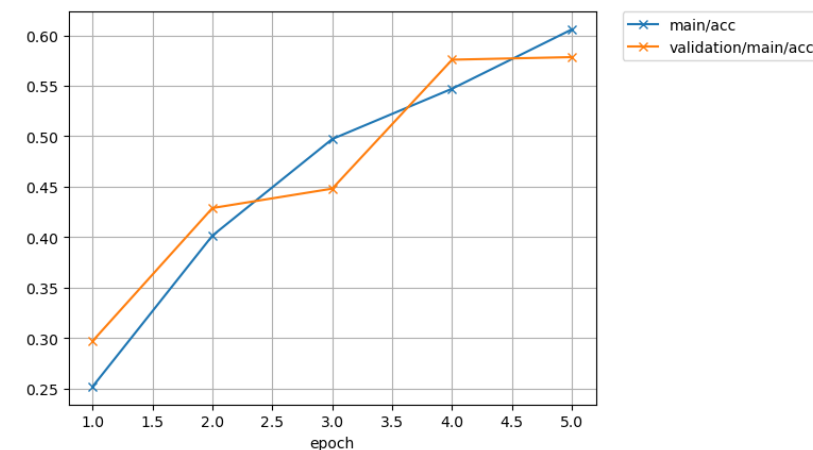
此 ASR module 用於辨識字母，設定了 28 個種類，執行到階段三時，將詞彙轉換成 JSON 格式儲存。



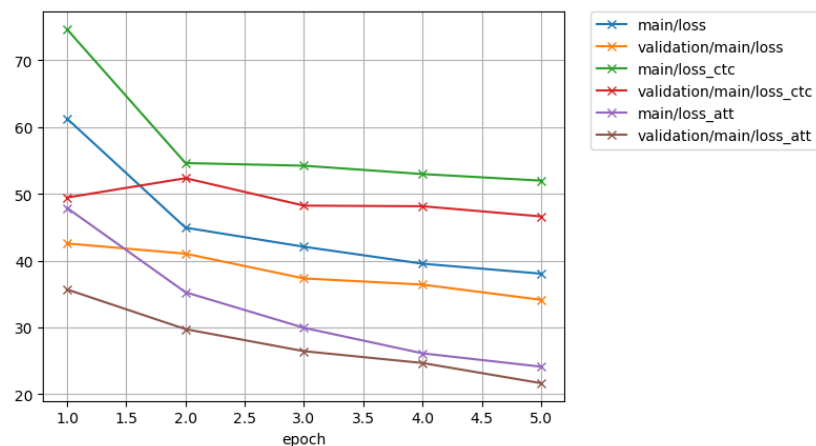
Try run the simple ASR model (an4) provided by ESPnet



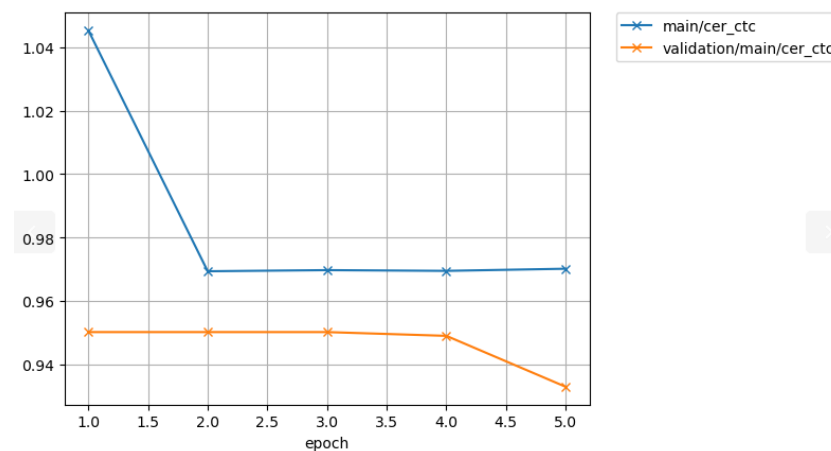
訓練結束後，會產出一些訓練數據，如圖一、二、三



(圖二) train/valid accuracy



(圖一) train/valid loss values



(圖三) train/valid character error rate