

VCC 2020 baseline design

Sian-Yi Chen

Advisor : Tay-Jyi Lin and Chingwei Yeh

Outline

- Action item

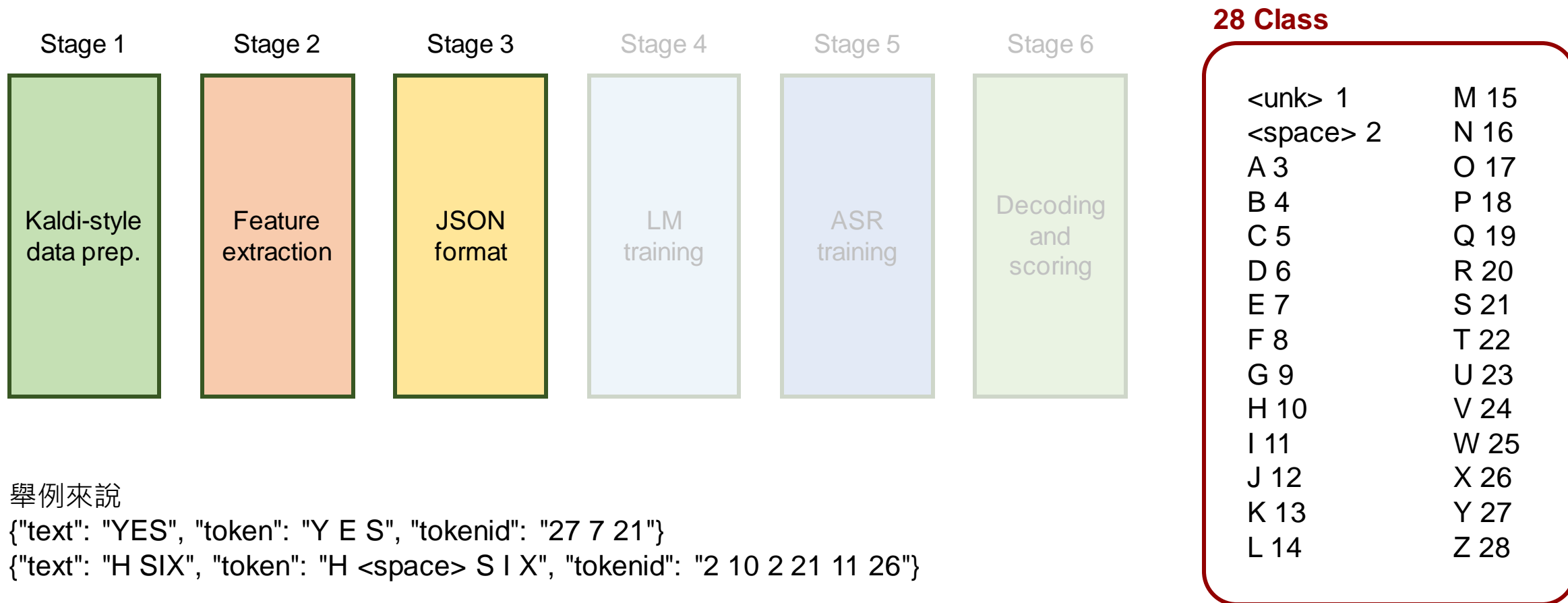
- 將 [1] 所提供的 open online resource 建構起來

- Status report

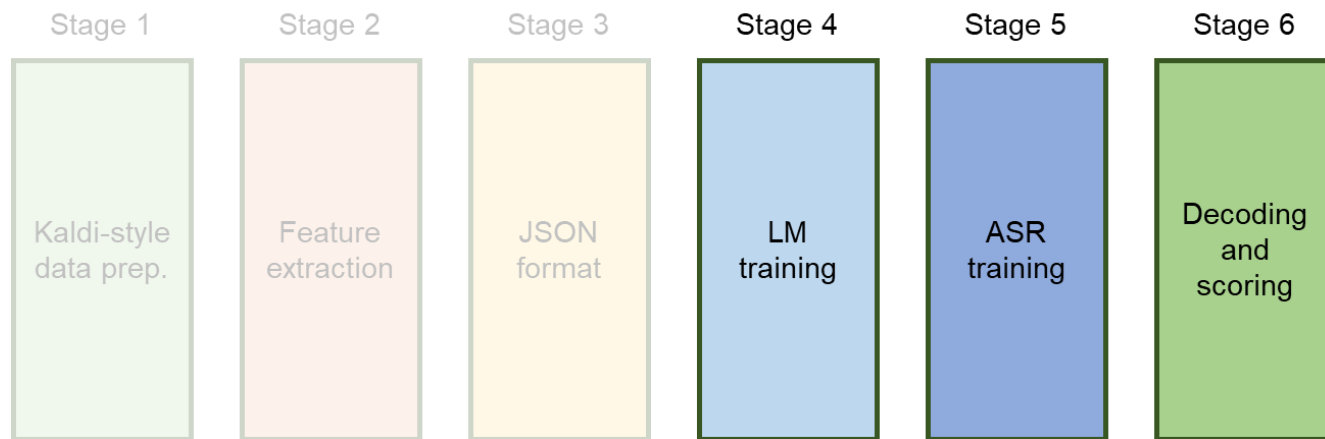
- 上週葉老師質疑我說 kaldi 需要特別多的資源很奇怪，我認為之前老師說電機的狀況應該是在有 GPU 的 server 上進行安裝，而一般 server 都會使用 linux 系統，但我們實驗室之前除了地下室的 powerAI 是使用 linux 系統之外，都是使用 windows 系統，而在 windows 系統下安裝 kaldi 就會變得困難許多。
 - 上週 Charlie 學長幫我重灌了 Linux 系統後，我設定好了 SSH 遠端設定，想說如果實驗室有其他想要使用 GPU 資源的可以連線進來使用。
 - 目前將需要的環境、套件、語料庫都已經處理好，但關於 ESPnet 提供的 ASR 功能還在研究要如何使用。
 - ESPnet 提供多達 34 種 ASR 任務。
 - 論文中使用的則是 LibriSpeech pretrained model (pytorch large Transformer with specaug (4 GPUs) + Large LSTM LM)
 - 而 ESPnet 數據顯示對於 LibriSpeech(en) 這個語料庫使用上方 ASR 模型：WER test-clean: 2.6%, test-other 5.7%
 - 在 ESPnet 提供的各種執行方法大致上皆由 6 個階段完成，因為資料層級複雜，因此先對於一個簡單的模型熟悉環境與操作流程，後面為簡單模型 (an4) 介紹與結果。
 - 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

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

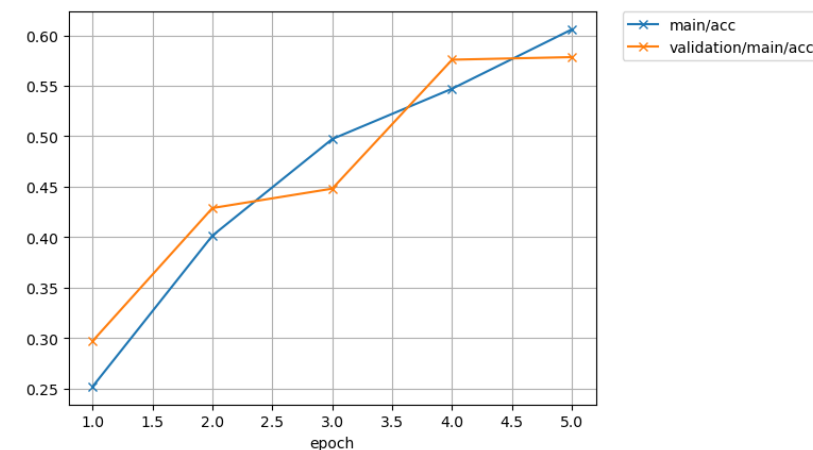
此 ASR module 用於辨識句子，設定了 28 個種類，執行到階段三時，將詞彙轉換成 JSON 格式儲存。



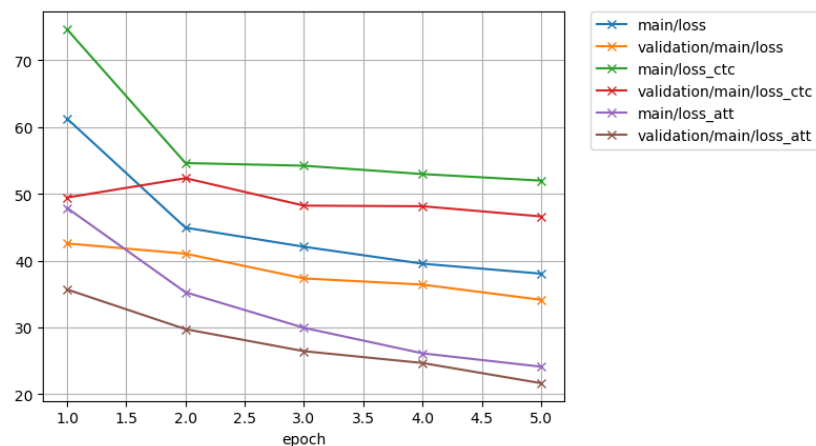
Try run the simple ASR model (an4) provided by ESPnet



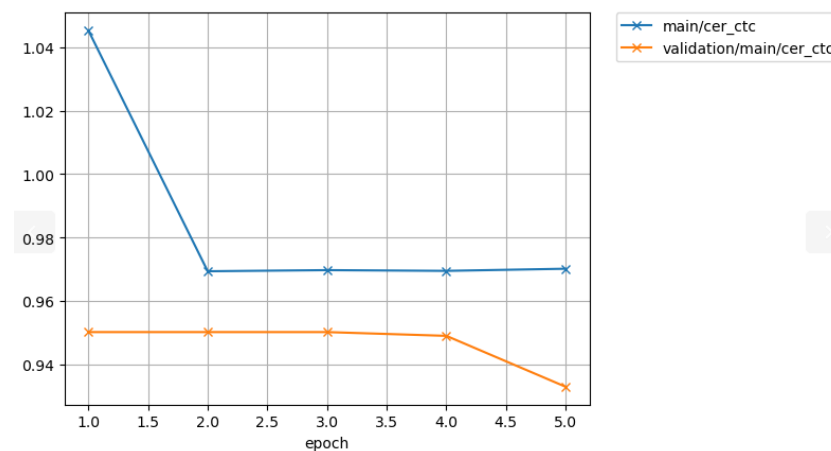
訓練結束後，會產出一些訓練數據，如圖一、二、三



(圖二) train/valid accuracy



(圖一) train/valid loss values



(圖三) train/valid character error rate