

SCIENTIFIC COMPUTING REPORT

THE MODELS YOU TRIED

- **Set-Model + Pitch-Model**

- 因為我們認為 onset、offset的預測和 pitch差很多，因此我們一開始先試著使用兩個 model，一個預測 onset和 offset，另一個預測 pitch。
- 在 training data上取得了不錯的效果，但 testing data上可能因為資料分布差異較大，反而效果變差許多。
- 後來發現直接用官方提供的 vocal.json裡頭的 pitch反而比自己預測得好，因此我們轉而修正我們的預測音符起始結束值的 LSTM_set，拿掉了 LSTM_pitch的 model，並取 vocal.json中 onset、offset範圍內的 pitch中位數，做為 note的 pitch。

- **Onset-Model + Offset-Model**

- 因為不確定 onset和 offset預測的性質有多相近，故我們也嘗試過用兩個 model去分開預測 onset和 offset，期望提升準確率。
- 結果顯示，準確率並沒有顯著提升，同時還會使訓練時間增加，因此我們最後仍只用一個model共同預測音符起始結束值。

- **Set-Model**

- 我們最後採用的版本只有使用單一一個 model，同時訓練 onset和 offset的預測。
- 最終版本的 model架構如下：
 - linear layer: 將資料維度提高到128維。
 - bidirectional encoder: 疊三個 layer，對前面 linear layer的輸出做雙向 encode。
 - fully connected layer: 將前面的 bidirectional encode結果作為輸入，輸出最終對 onset和 offset的機率預測。

DATASET PARTITIONING & CORRESPONDING PERFORMANCE EVALUATION

- **Dataset Partitioning**

- Training Data: 400
- Validation Data: 100
- Test Data: 1500

- **Performance Evaluation**

- F_measure: 加權 COn(20%)、COnP(60%)、COnPOff(20%)的 F_measure。

FEATURE NORMALIZATION, SELECTION, AND EXTRACTION

- **增加 Feature** 加上特定 feature 的二次項，作為新的 feature (ex. zero crossing rate, vocal_pitch...)。
- **Normalization** 用 standard scaler 去 fit training data 的 feature，並將其套用在 test data 的 feature 上。

IMPROVEMENT OVER THE BASELINE METHOD

Preprocess

- **修正Preprocess** 我們發現如果依照 sample_code 的作法產生 new label，會遇到前一個 note 的 offset 和現在的 note 的 onset 之間的時間差過短，導致前一個 frame 被標為前一個 note 的 offset 後，現在的 frame 和現在的 note 的 onset 之間的時間差已經大於 0.17，導致 onset 被漏標的情形。因此我們新增了一個機制：如果發現現在 frame 的時間已經大於現在的 note 的 onset，就直接將現在的 frame 標為新的 onset。
- **去除前奏和尾奏** 由於每首歌的 vocal.json 中都有頗長但沒什麼用的前奏和尾奏，故我們在 preprocessing 時，大幅去除 pitch 為零的前奏和尾奏，只在前後各保留五個 pitch 為零的 frame，這麼做使 training time 顯著降低。

Post-process

- **Onset & Offset 的挑選** train 好的 model 會輸出各個 frame 作為 onset (或 offset) 的機率。在進行 post-processing 的過程中，只要 frame 作為 onset (或 offset) 的機率超過我們所設的 threshold，就將其設為 onset 的候選人，如果它後面的 frame 是 onset 的機率低於現在這個 frame，則正式將該 frame 的時間標為現在的 note 的 onset。反之，若後面的 frame 機率更高，則捨棄現在的 frame，讓下一個 frame 成為 onset 的候選。
- **結合 PV2Note** 在過程中，我們發現我們所預測出的 answer.json，其中的資料筆數比 ground truth 的資料筆數少很多。故我們參考了 PV2Note 中的概念，在對 label 結果做 post_processing 時，如果現在 frame 的 pitch 和目前 onset 的 pitch，兩者間差值超過某個 threshold，就將現在的 frame 分到新的 note (也就是說，前一個 frame 會成為前一個 note 的 offset，現在的 frame 則是新的 note 的 onset)。

INSIGHT AND CONCLUSIONS

一開始覺得大概一定要用 Machine Learning 才会有好結果，但後來詢問沒使用 Machine Learning 的組別後，發現兩者結果不相上下，因此決定加上一點 Domain Knowledge。沒想到這麼做後，performance 的提升還蠻顯著的。由此可見，在做預測的時候，Machine Learning 技術和 Domain Knowledge 相輔相承，或許能發揮一加一大於二的效果。

DIVISION OF LABOR

- B05902010 資工四 張頌平：ML method and Rule-base method(40%)
- B06902047 資工三 陳彥：Rule-base method and Report(30%)
- B06902135 資工三 蔡宜倫：Rule-base method and Report(30%)