Tomofun 狗音辨識 AI 百萬挑戰賽

初賽賽後報告

[建議內容長度 2 ~ 5 頁，文件需提供足夠資訊讓主辦單位得以審查並重現成果，若繳交文件不符合規格，主辦單位保有權力得以取消晉級資格。]

隊伍：哈哈哈

成員：姜立垣

1. 環境

在 Colab 上

OS:

Ubuntu 18.04.5 LTS

Language:

python

Packages:

torch==1.9.0

librosa==0.8.1

warmup-scheduler==0.3.2

audiomentations==0.16.0

Pretrained Model:

PANN(https://zenodo.org/record/3987831/files/Cnn14\_mAP%3D0.431.pth?download=1)

1. 作法摘要與介紹

主要使用預訓練模型再於 public train data 上 finetune，cv 可達 0.9759，Private LB 可達 0.9785。

1. 資料處理

各項參數設置如下：

hop\_length = 256

fmin = 0

fmax = 4000

n\_mels = 64

n\_fft = 1024

padmode = 'reflect'

res\_type = 'kaiser\_best'

Data Augmentation:

AddGaussianNoise(min\_amplitude=0.001, max\_amplitude=0.015, p=0.3)

AddGaussianSNR(p=0.3)

Gain(min\_gain\_in\_db=-12, max\_gain\_in\_db=12, p=0.3)

PitchShift(min\_semitones=-4, max\_semitones=4, p=0.3)

TimeStretch(min\_rate=0.8, max\_rate=1.2, p=0.3)

Shift(min\_fraction=-0.5, max\_fraction=0.5, rollover=False, p=0.3)

主要是增加雜訊與變化音調、時間拉伸、偏移，但是不希望失真太多，所以不會過度擴增，且 transform 機率也調低

1. 模型架構

PANN model(https://github.com/qiuqiangkong/audioset\_tagging\_cnn)

Qiuqiang Kong, Yin Cao, Turab Iqbal, Yuxuan Wang, Wenwu Wang, Mark D. Plumbley. "PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition." arXiv preprint arXiv:1912.10211 (2019).

1. 訓練方式

StratifiedKFold 切 5 份，取其中四分作為訓練，一份為驗證

總共 200 epoch (warmup\_epoch=20 + cosine\_epoch=180)

warmup\_factor = 10

batch\_size = 128

LR = 5e-4

Optimizer: Adam

Scheduler: GradualWarmupScheduler + CosineAnnealingLR

1. 實驗分析與結論

說明此模型在訓練/驗證/測試集上的表現，選擇此模型的好處，以及未來可再改進的方向。

此模型表現：訓練集0.9847，驗證集 0.9912，測試集 0.9759。

選擇預訓練模型好在已有前人建置好的基礎，我們再加以微調基本上可取得不錯成果。未來將嘗試 metric learning 以更好用於區分 others，以及 self supervised learning 模型，甚至可使用 public test 與 private test 資料加以訓練 embedding 最後用於分類。

1. 程式碼

於 colab 開啟 GPU 後直接執行 ipynb 即可

1. 使用的外部資源與參考文獻

https://github.com/ryanwongsa/kaggle-birdsong-recognition/tree/4ad1aa4ed99bc097289c7593c55bc09234e0fc59

1. 心得感想與建議

謝謝主辦單位用心舉辦，希望以後還有機會參加！