Tomofun 狗音辨識 AI 百萬挑戰賽

初賽賽後報告

隊伍：無法顯示

成員：沈之謙、紀宥紘

1. 環境

作業系統:window10

語言:python

套件:

Tensorflow1.14.0, keras2.3.1, audiomentations0.16.0,

預訓練模型:無

額外資料集: 50種環境聲音分類使用雨聲、人聲等作為背景音(https://god.yanxishe.com/37)

1. 作法摘要與介紹

使用5秒8000Hz的音頻檔案，轉換為128\*512的頻譜圖，並使用CNN模型(約400K參數)進行訓練，訓練集部分使用資料增強(包含時間平移、隨機遮罩、添加背景噪音等)，最終LB:0.987

訓練史:

128\*128的頻譜圖+簡單模型-> LB:0.971

128\*128的頻譜圖+加深模型-> LB: 0.977

128\*512的頻譜圖-> LB: 0.982

Fmax設定為4000-> LB: 0.983

資料增強-> LB: 0.985

資料增強+背景音-> LB: 0.987

1. 資料處理

將音頻資料轉為頻譜圖進行CNN訓練。

頻譜圖相關參數說明:

頻譜圖長寬: 128\*512

最初設定為128\*128，後來發現某些分類不明顯，便將大小改為128\*512，將寬度加寬，讓時間細節顯現。

最高頻率= 4000

因音頻資料採樣率為8000Hz，因此將最高頻率設定為 4000。

其他參數無特別修改，詳細內容如程式碼內容。

1. 模型架構

模型使用了BatchNormalization，並使用relu激活函數。

在底層使用GlobalAveragePooling2D，取代Flatten才不會導致輸出參數過多。

因訓練模型時，出現極度的overfitting(train與val差異極大)，因此在模型中大量使用Dropout、SpatialDropout2D。

過程中也常是許多其他的模型，但效果看起來大同小異，因此並無花費太多時間在模型調整上。

1. 訓練方式

將資料切割為五份，分別訓練5個模型，用五個模型之平均作為結果。

分成兩次訓練，減少花費時間。

1.pre\_train

使用原始的訓練集資料，訓練約500個epoch，並且使用ModelCheckpoint，保留每10個epoch之結果。

2.aug\_train

使用資料增強之訓練集資料(包含時間平移、隨機遮罩、添加背景噪音等)，訓練約50個epoch。

使成績LB0.983 -> LB0.987

1. 實驗分析與結論

1.在訓練集上的數據大多為乾淨數據，大多沒有背景音，因此無聲的頻譜圖呈現黑色，但在測試集頻譜圖上，大多都是包含雜音的，因此用資料增強解決此問題。

2.在測試集上具有訓練集沒有出現的聲音，像是車聲(應該是歸在其他類)，因已在資料增強中新增了許多有可能會遇到的聲音當背景音。

3.其他類處理:想使用sigmoid取代softmax，才不會把分類強迫分在某一類，可以定義如果都沒有達到門檻，則定義為其他類，但目前還沒有嘗試出好的方法。

1. 程式碼

如附檔

1. 使用的外部資源與參考文獻

50種環境聲音分類 (<https://god.yanxishe.com/37>)

1. 心得感想與建議

測試集與訓練集具有顯著差異

目前不太確定決賽的準備方向