GSCD 之單字 corpus 數量與 accuracy 分析及 SpecAugment 效益

Student : 陳憲億、胡祐嘉
Advisor : Chingwei Yeh and Tay-Jyi Lin

Outline

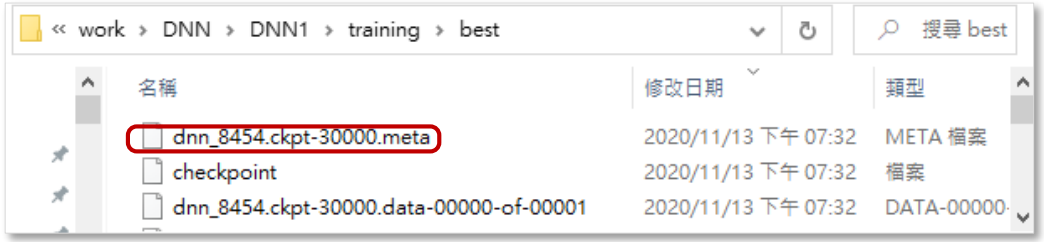
Action item：設計實驗證明 SpecAugment 方法是有成效的，並在 outline 部分呈現 7 成內容

- **實驗目的：**證明 SpecAugment 方法有用
- **實驗方法 (流程)：**
  1. 將 GSCD 12 種關鍵字中的 “yes” 語料從 4000 筆開始下降至 250 筆 (將現有的語料隨機刪除至需求筆數)，並依序將 GSCD 放入 Hello edge 中做訓練，訓練結束後，使用原始的 GSCD 做測試，最後由產生出來的混淆矩陣來計算每次 yes 的辨識率，藉以找出語料數量為多少時，將不足以訓練出理想值(目前進度)
  2. 找到正確率嚴重下降的點後，使用 SpecAugment 將不足的語料增量到與原始語料相似數量 (≈ 4000)，觀察 yes 辨識率是否有效上升
- **實驗參數：**
  1. 使用 12 種關鍵字 (silence, unknow, yes, no, up, down, left, right, on, off, stop, go)
  2. 使用 dnn 模型，大小為 144 x 144 x 144
  3. learning rate： 0.0005, 0.0001, 0.00002
  4. training step： 10000, 10000, 10000
- **實驗猜想：**當訓練語料不足時，辨識率產生嚴重下降，此時使用 SpecAugment 將語料做增量，辨識率可以有效上升
- **實驗結果：**目前進行到第一步驟，發現語料衰減到 250 筆時，原先約為 85% 的辨識率下跌至 47% 左右
- **下周規劃：**在模型數量 1000 筆到 250 筆之間，將減量單位變小，找出平滑曲線，並進行第二步驟

實驗進行流程

1 輸入指令： (spec) D:\Work_Space\KeyWordSpotting-for-MCU-master>python train.py --model_architecture dnn --model_size_info 144 144 144 --dct_coefficient_count 10 --window_size_ms 40 --window_stride_ms 40 --learning_rate 0.0005,0.0001,0.00002 --how_many_training_steps 10000,10000,10000 --summaries_dir work/DNN/DNN1/retrain_logs --train_dir work/DNN/DNN1/training --data_url="" --data_dir="./speech_dataset/"

使用複製的GSCD語料，保留原始資料做測試



執行完成，訓練數據以 checkpoint 的方式保存

2 輸入指令： (spec) D:\Work_Space\KeyWordSpotting-for-MCU-master>python test.py --model_architecture dnn --model_size_info 144 144 144 --dct_coefficient_count 10 --window_size_ms 40 --window_stride_ms 40 --checkpoint="./work/DNN/DNN1/training/best/dnn_8454.ckpt-30000" --data_url=""

```
I1114 21:14:28.288953 15500 test.py:112] Confusion Matrix:
 [[3121    0    0    0    0    0    0    0    0    0    0    0]
 [   2 2261   46   49   73  126   76  104   84   33   63  132]
 [   2   74 2801   35    3   29  143    5    2   21    7   12]
 [   1   58    9 2764   17  106   15   13    5    2   26  210]
 [   0   38    1   14 2592   18   11   20   35  123  115   37]
 [   0   56   16  104   23 2730   26    6   37    5   25   53]
 [   0   52   63   12   35   18 2709   74   10   16   10   21]
 [   0   93    1    6   15   13   81 2671   18    3    1   10]
 [   3   49    0    3   69   20    5   10 2784   81   25   11]
 [   7   22    1    2  119    4   12    9   63 2751   33   22]
 [   5   32    3    7  102   14   19    3    2   16 2896   21]
 [   2   75    6  314   63  133   20    7   11   19   36 2413]]
INFO:tensorflow:Training accuracy = 88.13% (N=36871)
I1114 21:14:28.293978 15500 test.py:114] Training accuracy = 88.13% (N=36871)
```

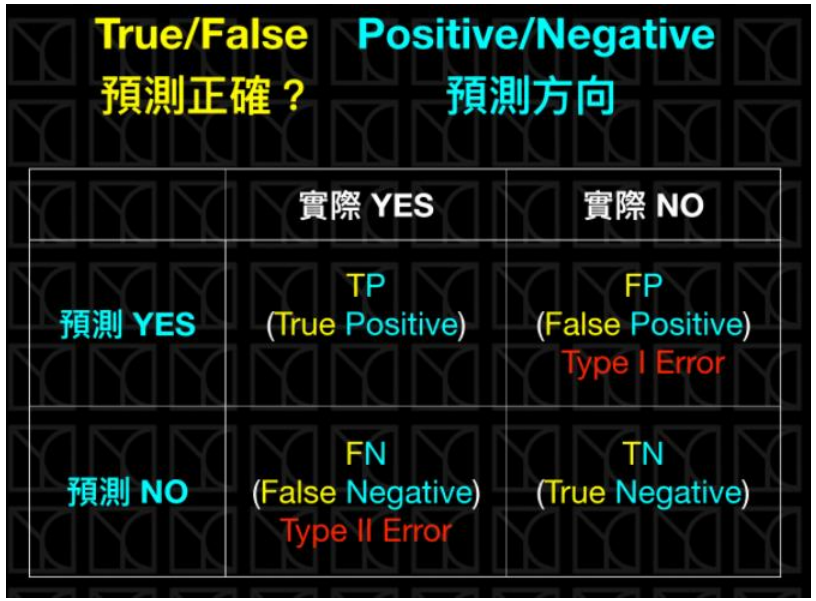
執行完成，產生 Training、Test 的混淆矩陣

3 GSCD語料處置： D:\Work_Space\KeyWordSpotting-for-MCU-master\speech_dataset (遞減組，用於訓練的語料)

D:\Work_Space\Google_speech_dataset (保持不變組，用於測試的語料)

混淆矩陣(confusion matrix)

混淆矩陣(confusion matrix)



Accuracy = (TP + TN) / total N

計算準確率：

```
INFO:tensorflow:set_size=5032
INFO:tensorflow:Confusion Matrix:
          silence unknow yes  no   up  down left right on  off stop go
silence  [[420    0    0    0    0    0    0    0    0    0    0    0]
unknow   [  0  299   13   11   10   22    4   11   16    8   12   14]
yes      [  0   13  491    7    0    4   17    0    0    2    0    3]
no       [  0    3    8  327    0   20    6    1    0    1    0   39]
up       [  1    9    1    4  363    1    4    5    8   11   12    6]
down     [  0   12    7   30    3  322    4    3    6    1    4   14]
left     [  0   10   32    3    7    3  334   17    0    3    2    1]
right    [  0   18    3    2    3    0   13  347    5    3    0    2]
on       [  0   14    0    1    7   11    1    0  340   16    3    3]
off      [  1    6    1    3   21    0    6    4   16  336    3    5]
stop     [  0    4    1    2   19    4    4    1    2    3  365    6]
go       [  0   15    6   70    9   11    4    0    0    1    1  285]]
INFO:tensorflow:Test accuracy = 84.04% (N=5032)
```

$$\text{Yes Accuracy} = \frac{491}{13 + 491 + 7 + 4 + 17 + 2 + 3} = 91.43\%$$

Total Accuracy = (TP + TN) / total N
= 4229 / 5032
= 84.04%

Accuracy 數據分析

類別 \ 數量(筆)	4000	2500	2000	1500	1000	500	250
Training Accuracy (12個 calss 平均)	87.16%	86.59%	88.24%	87.60%	87.15%	86.75%	85.10%
Test Accuracy (12個 calss 平均)	83.96%	83.25%	83.64%	83.13%	82.65%	82.63%	80.30%
Training Accuracy (yes音檔)	93.45%	85.17%	89.37%	84.86%	80.65%	71.64%	53.99%
Test Accuracy (yes音檔)	91.06%	84.29%	88.64%	84.05%	80.67%	72.70%	47.82%