

Introduction to Digital Speech Processing – Homework2

R07922106 曾俊為

一開始我將原檔跑過一次，accuracy 為 75.30，確認環境與初始值的設定正確。

```
(kaldi) root@cf3d3de75d28a:/opt/kaldi/dsp-hw2-1# bash 4-test.sh
Converting acoustic models to HTK format
viterbi/mono/final.mmf viterbi/mono/tiedlist exist , skipping ...
Generating results for test set with acoustic weight = [ 0.87 ]
output -> viterbi/mono/test.mlf
log -> viterbi/mono/log/latgen.test.log
result -> viterbi/mono/test.rec
accuracy -> [ 75.30 ] %

Execution time for whole script = 00 hours 00 mins 03 secs
```

再來我先從 train.sh 下手，一開始我去改變了 Gaussian 的數量，將 initial Gaussian 的數量固定之後，慢慢將最終的 Gaussian 數量由 100 調升到 1000，由結果得知，accuracy 有隨著 Gaussian 數量增加而上升，但是效果很有限。Accuracy 大約到了 80.5%。(下圖 table1)

totgauss	accuracy
100	79.39
200	79.22
300	80.02
500	80.08
1000	80.48

(table1)

numiters	accuracy
6	83.30
7	86.82
8	87.68
10	89.18
15	90.73

(table2)

所以接下來，我從 iteration 的數量下手，從初始值 5 慢慢調升到 15，maxiterinc 也要跟著做些改動，讓 last iteration 有到達 totalGaussian 的數量。由結果可以發現 iteration 的改動對 accuracy 有滿不錯的提升。

不過隨著 iteration 的提高，training 的時間會越來越久(光 training 就需要 3 分半加上其他流程就可能超過 4 分鐘了)，而且 accuracy 的提升也漸漸趨緩，加上擔心 iteration 太多會有 overfitting 的情況，所以最後選擇 iteration 在 15 的位置，accuracy 來到大約 90.7%。(上圖 table2)

再來我則是去改變 test.sh 中 opt_acwt 的值，也就是 acoustic model 占的比重，初始值為 0.87，我先稍微往上調到 0.9 後，接下來一路往下調到 0.1。(下圖)

opt_acwt	accuracy
0.9	90.56
0.8	91.08
0.7	92.11
0.5	93.49
0.3	95.22
0.2	95.74
0.1	95.16

最後發現 acoustic model 在 decode 時占大約 20%時，accuracy 會最高。

另外我有嘗試去更改 topo.proto 裡面 states 的數量和 transition 的值，但是依然沒辦法到達 99%以上，且常常會有 error 出現 (可能是我 transition 的設定不對所致)。

綜合以上的步驟，最終我選擇使用 iteration 為 15，Gaussians 總數為 1000，且 acoustic model 占 20%，除了 3-train.sh 和 4-test.sh 之外，其餘檔案的參數皆沒改變的 model，得到的 accuracy 為 95.74。

```
(kaldi) root@cf3d3de75d28a:/opt/kaldi/dsp-hw2-1# bash 4-test.sh
Converting acoustic models to HTK format
viterbi/mono/final.mmf viterbi/mono/tiedlist exist , skipping ...
Generating results for test set with acoustic weight = [ 0.2 ]
output -> viterbi/mono/test.mlf
log -> viterbi/mono/log/latgen.test.log
result -> viterbi/mono/test.rec
accuracy -> [ 95.74 ] %

Execution time for whole script = 00 hours 00 mins 05 secs
```