MSc Electronic and Computer Engineering

Data Mining and Machine Learning (2019)

Lab 3 – Speech recognition by using HTK

Group 12

Miaoyu Niu (1893824)

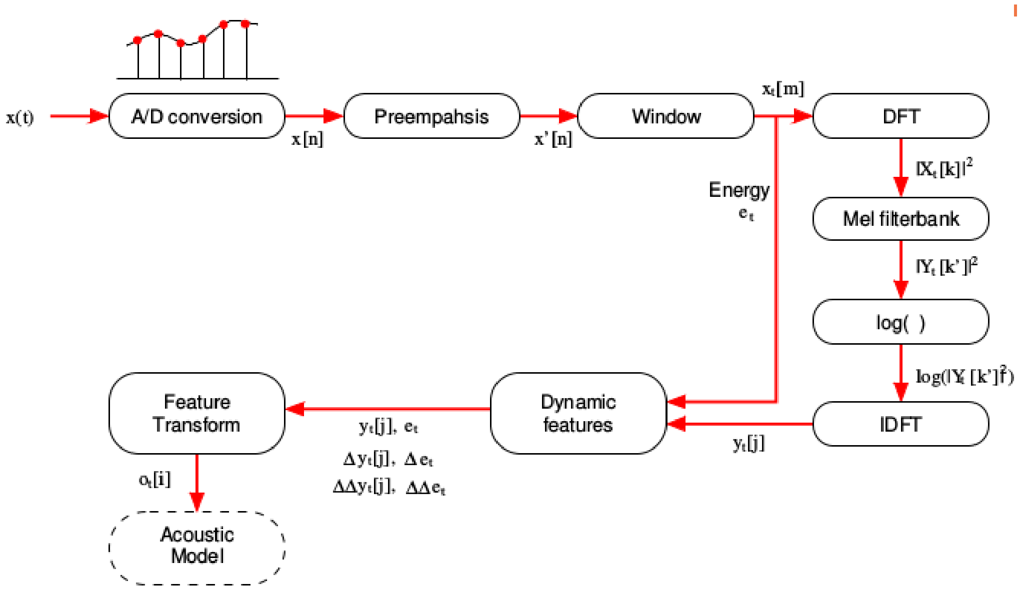
Kai Chen (1948361)

2019-3-6

UNIVERSITY OF BIRMINGHAM

Task1

Before starting this lab, I checked some knowledge about MFCC firstly, as shown below, maybe it will not help me in this lab, but it can help me know what is MFCC.



This is pre-process of MFCC. The original speech signal is digitally converted by A/D conversion, and the high frequency component is boosted by pre-emphasis, followed by windowing, and the processed signal is processed in two aspects. One aspect is to extract the cepstrum feature, that is, after the discrete Fourier After the transformation, the spectrum amplitude is squared, passed through the Mel filter bank, and then logarithmically transformed, and finally the inverse of the discrete Fourier transform is used to obtain the cepstrum feature; on the other hand, the energy of the signal after windowing is obtained. The aspects are combined to form a dynamic feature, and finally the feature transformation is performed to obtain an acoustic model.

Code:

```
F:\lab3\LabASR>HTK3.2bin\\HCompV -C config/config_train_MFCC_E -o hmmdef -f 0.01 -m -S list/listTrainFullPath_LabDMML_CLEAN1.scp -M hmmsTrained/hmm0 lib/proto_s1d13_st8m1_LabDMML_MFCC_E

F:\lab3\LabASR>HTK3.2bin\\models_1mixsil hmmsTrained/hmm0/hmmdef hmmsTrained/hmm0/models
CREATING SILENCE MODEL

F:\lab3\LabASR>HTK3.2bin\\macro 13 MFCC_E hmmsTrained/hmm0/vFloors hmmsTrained/hmm0/macros
```

```
F:\lab3\LabASR>HTK3.2bin\\HERest -C config/config_train_MFCC_E -I label/label_LabDMML_noSP.mlf -t 250.0 150.0 1000.0 -S list/listTrainFullPath_LabDMML_CLEAN1.scp -H hmmsTrained/hmm0/macros -H hmmsTrained/hmm0/models -M hmmsTrained/hmm1 lib/wordList_noSP
Pruning-On[250.0 150.0 1000.0]

F:\lab3\LabASR>HTK3.2bin\\HERest -C config/config_train_MFCC_E -I label/label_LabDMML_noSP.mlf -t 250.0 150.0 1000.0 -S list/listTrainFullPath_LabDMML_CLEAN1.scp -H hmmsTrained/hmm1/macros -H hmmsTrained/hmm1/models -M hmmsTrained/hmm2 lib/wordList_noSP
Pruning-On[250.0 150.0 1000.0]

F:\lab3\LabASR>HTK3.2bin\\HERest -C config/config_train_MFCC_E -I label/label_LabDMML_noSP.mlf -t 250.0 150.0 1000.0 -S list/listTrainFullPath_LabDMML_CLEAN1.scp -H hmmsTrained/hmm2/macros -H hmmsTrained/hmm2/models -M hmmsTrained/hmm3 lib/wordList_noSP
Pruning-On[250.0 150.0 1000.0]
```

```
F:\lab3\LabASR>HTK3.2bin\\spmodel_gen hmmsTrained/hmm3/models hmmsTrained/hmm4/models

F:\lab3\LabASR>HTK3.2bin\\HHEd -H hmmsTrained/hmm4/macros -H hmmsTrained/hmm4/models -M hmmsTrained/hmm5 lib/tieSIL
andSP_LabDMML.hed lib/wordList_withSP

F:\lab3\LabASR>HTK3.2bin\\HERest -C config/config_train_MFCC_E -I label/label_LabDMML_withSP.mlf -t 250.0 150.0 100
0.0 -S list/listTrainFullPath_LabDMML_CLEAN1.scp -H hmmsTrained/hmm5/macros -H hmmsTrained/hmm5/models -M hmmsTrain
ed/hmm6 lib/wordList_withSP
Pruning-On[250.0 150.0 1000.0]

F:\lab3\LabASR>HTK3.2bin\\HERest -C config/config_train_MFCC_E -I label/label_LabDMML_withSP.mlf -t 250.0 150.0 100
0.0 -S list/listTrainFullPath_LabDMML_CLEAN1.scp -H hmmsTrained/hmm6/macros -H hmmsTrained/hmm6/models -M hmmsTrain
ed/hmm7 lib/wordList_withSP
Pruning-On[250.0 150.0 1000.0]

F:\lab3\LabASR>HTK3.2bin\\HERest -C config/config_train_MFCC_E -I label/label_LabDMML_withSP.mlf -t 250.0 150.0 100
0.0 -S list/listTrainFullPath_LabDMML_CLEAN1.scp -H hmmsTrained/hmm7/macros -H hmmsTrained/hmm7/models -M hmmsTrain
ed/hmm8 lib/wordList_withSP
Pruning-On[250.0 150.0 1000.0]
```

```
F:\lab3\LabASR>HTK3.2bin\\HVite -H hmmsTrained/hmm8/macros -H hmmsTrained/hmm8/models -S list/listTestFullPath_LabDMML_CLEA
N1.scp -C config/config_test_MFCC_E -w lib/wordNetwork -i result/result.mlf -p 10 -s 0.0 lib/wordDict lib/wordList_withSP
Building word sil
Building word sp
Building word oh
Building word !NULL
Building word zero
Building word nine
Building word eight
Building word seven
Building word six
Building word five
Building word four
Building word three
Building word two
Building word one
Building word sil
Building word !NULL
Building word !NULL
33 nodes
 = 17 word (0 null), init 14, int 0, fin 0, cf 0
78 links = int 14, ext 62 (0 for null words)
```

Above codes are written when I train the tutorial parts.

In command window, I enter flowing commands:

F:\lab3\LabASR>C:\Perl\bin\perl ASR_LabDMML_MFCC_E.pl

F:\lab3\LabASR>C:\Perl\bin\perl ASR_LabDMML_onlyTest_MFCC_E.pl

```
F:\lab3\LabASR>C:\Perl\bin\perl ASR_LabDMML_MFCC_E.pl

F:\lab3\LabASR>C:\Perl\bin\perl ASR_LabDMML_onlyTest_MFCC_E.pl
```

Result:

```
------------------------ Overall Results --------------------------
SENT: %Correct=83.85 [H=161, S=31, N=192]
WORD: %Corr=98.15, Acc=95.72 [H=849, D=0, S=16, I=21, N=865]
===================================================================
```

Analysis of result:

In this task, Perl scripts automate the training and testing process. By modifying some of the parameters in the recognition system, it not only has static MFCC features, but also delta and delta-delta features. The training process uses clean test data. From result, we can know that there are no deletion errors, 16 replacement errors and 21 insertion errors. It has 865 words totally, 849 words were accepted which occupies 98.15%. According to the result, by calculating the MFCC trajectories and adding delta and delta-delta features to the original features can improve the performance of speech recognition.

Task2

Code:

F:\lab3\LabASR>C:\Perl\bin\perl ASR_LabDMML_3mix_MFCC_E.pl
F:\lab3\LabASR>C:\Perl\bin\perl ASR_LabDMML_onlyTest_3mix_MFCC_E.pl

```
F:\lab3\LabASR>C:\Perl\bin\perl ASR_LabDMML_3mix_MFCC_E.pl

F:\lab3\LabASR>C:\Perl\bin\perl ASR_LabDMML_onlyTest_3mix_MFCC_E.pl
```

Result:

```
------------------------ Overall Results --------------------------
SENT: %Correct=89.06 [H=171, S=21, N=192]
WORD: %Corr=98.84, Acc=96.99 [H=855, D=0, S=10, I=16, N=865]
===================================================================
```

Analysis of result:

In this task, I use another Perl script to develop a recognition system that uses the MFCC_E_D_A features and uses 3 Gaussian mixture components in each state. From the result, I can know that it has 865 words in total, and system can recognize 855 words which occupies 98.84%. There are no deletion errors, exit 10 substitution errors and 16 insertion errors. We can know that the accuracy figure is improved from 95.72% to 96.99% and the percentage correct increases from 98.15% to 98.84%.

Task3

In this task, We modify the value of p in each testing in order to balance the amount of deletion and insertion errors. Besides, using clean data.

Result:

Clean data (P=-40.5)

```
------------------------ Overall Results --------------------------
SENT: %Correct=91.67 [H=176, S=16, N=192]
WORD: %Corr=98.38, Acc=98.03 [H=851, D=3, S=11, I=3, N=865]
===================================================================
```

First sort of noisy data (N1_SNR10)

```
F:\lab3\LabASR>HTK3.2bin\\HCopy -C config/config_HCopy_MFCC_E -S list/listTestHCopy_LabDMML_N1_SNR10.scp

F:\lab3\LabASR>C:\Perl\bin\perl ASR_LabDMML_3mix_MFCC_E.pl

F:\lab3\LabASR>C:\Perl\bin\perl ASR_LabDMML_onlyTest_3mix_MFCC_E.pl
```

Turning p value to -87

```
------------------------ Overall Results --------------------------
SENT: %Correct=13.02 [H=25, S=167, N=192]
WORD: %Corr=67.28, Acc=59.08 [H=582, D=72, S=211, I=71, N=865]
===================================================================
```

Second sort of noisy data (N2_SNR15)

```
F:\lab3\LabASR>HTK3.2bin\\HCopy -C config/config_HCopy_MFCC_E -S list/listTestHCopy_LabDMML_N1_SNR15.scp

F:\lab3\LabASR>C:\Perl\bin\perl ASR_LabDMML_onlyTest_3mix_MFCC_E.pl
```

Turning p value to -101

```
------------------------ Overall Results --------------------------
SENT: %Correct=51.04 [H=98, S=94, N=192]
WORD: %Corr=86.24, Acc=82.66 [H=746, D=31, S=88, I=31, N=865]
===================================================================
```

Third sort of noisy data (N2_SNR20)

```
F:\lab3\LabASR>HTK3.2bin\\HCopy -C config/config_HCopy_MFCC_E -S list/listTestHCopy_LabDMML_N1_SNR20.scp
F:\lab3\LabASR>C:\Perl\bin\perl ASR_LabDMML_onlyTest_3mix_MFCC_E.pl
```

Turning p value to -60.5

```
------------------------ Overall Results --------------------------
SENT: %Correct=68.23 [H=131, S=61, N=192]
WORD: %Corr=93.06, Acc=90.98 [H=805, D=18, S=42, I=18, N=865]
===================================================================
```

Analysis of result:

By observing the difference of the results tested on three different noisy data, we can find that the accuracy and correctness are improved with the decrease of the value of deletion error, substitution error and insertion error. The accuracy of the experiment tested on noisy data changes from 59.08% to 82.66% then to 90.98% with deletion error changes from 72 to 31 then to 18, substitution error changes from 211 to 88 then to 42 and insertion error changes from 71 to 31 then to 18. Although the accuracy and correctness has improved a lot by testing on the third kind of noisy data, they are still not as good as those tested on clean data. We can make a conclusion that the values of accuracy and correctness tested on noisy data are less than the result tested on clean data and they can be improved with the reduction of the insertion errors.

Task4

Result:

Using the combined set of clean and all noisy training data:

```
------------------------ Overall Results --------------------------
SENT: %Correct=88.54 [H=170, S=22, N=192]
WORD: %Corr=99.08, Acc=96.99 [H=857, D=0, S=8, I=18, N=865]
===================================================================
```

Turning p value to -75

```
------------------------ Overall Results --------------------------
SENT: %Correct=93.23 [H=179, S=13, N=192]
WORD: %Corr=98.61, Acc=98.15 [H=853, D=4, S=8, I=4, N=865]
===================================================================
```

First sort of noisy data (N2_SNR10)

Turning p value to -17.5

```
------------------------ Overall Results --------------------------
SENT: %Correct=78.65 [H=151, S=41, N=192]
WORD: %Corr=94.80, Acc=93.99 [H=820, D=7, S=38, I=7, N=865]
===================================================================
```

Second sort of noisy data (N2_SNR15)

Turning p value to -32

```
------------------------ Overall Results --------------------------
SENT: %Correct=90.10 [H=173, S=19, N=192]
WORD: %Corr=97.69, Acc=97.34 [H=845, D=3, S=17, I=3, N=865]
===================================================================
```

Third sort of noisy data (N2_SNR20)

Turning p value to -45.5

```
----------------------- Overall Results -------------------------
SENT: %Correct=90.10 [H=173, S=19, N=192]
WORD: %Corr=97.92, Acc=97.46 [H=847, D=4, S=14, I=4, N=865]
================================================================
```

Analysis of result:

In this task, we need to train the system by using the combined set of clean and all three kinds of noisy training data. The experimental evaluation of this system is performed separately on clean and each noisy testing data. By observing the result of the experiments tested on three kinds of different noisy data, we can find that the accuracy and correctness of the third kind of noisy data are the best. When testing the first kind of noise, the result of accuracy is 93.9%, compared with 59.08% in task 3. The result is the same for the second and third kind of noise.