音声認識システムにおける
認識率の評価

小林達矢

2020年2月10日

目 次

1　始めに

Hey,Siri! OK,google など私たちの身の回りには私たちの音声によって起動する機器がたくさん存在する。そこで実際に学習させて認識させるために今回 HTK を用いて実験した。

2　フリーソフトを使った音声認識

2.1　音声認識に用いたフリーソフト

今回音声認識をやるにあたって主に 3 つのフリーソフトを使用したのでそれらを順に説明する。一つ目は iPhone に搭載されているボイスレコーダーである。このアプリを用いることで録音した音声を wave ファイ r つとして保存することができる。二つ目は wavesurfer である。これはボイスレコーダーで保存した wave ファイルをラベル付けするために使用した。(ラベル付けに関したは後の説で説明する) 三つ目は、HTK(Hidden Model Toolkit) である。これはケンブリッジ大学で開発され音声認識する際に利用する。これは、教育・研究目的ならば無料で使用することができる。

2.2　HTK を用いた音声認識

2.2.1　音声の内容

今回は数字の 0 から 9 を録音して正しく音声が認識されているかを確認する。また今回はそれぞれ 3 回ずつ録音して実験している。

2.2.2　音声の録音とラベル付け

iPhone のボイスレコーダーを用いて wave ファイルの形式で保存する。その保存した wave ファイルを wavesurefer に取り込む。この時の音声は発生している音の前後に無音時間があるので、その無音のとこを sil(silent の略)、発声している部分をそれぞれの数字でラベル付けをする。

2.2.3 特徴抽出

Hcopy コマンドによる特徴抽出を行う。HCopy コマンドは

HCopy

```
HCopy [オプション]　音声ファイル名　特徴ファイル名
  C 構成ファイル名を指定
  S スクリプトファイル名を指定
```

の形式で扱う。
実際には HCopy -C config.hcopy -S script.hcopy
として使用した。今回使用する音声ファイル名を config.hcopy とし、内容は以下のものである。

config.hcopy

```
SOURCEFORMAT = WAV
SOURCEKIND = WAVEFORM
SOURCERATE = 625
TARGETKIND = MFCC_0_D_A
TARGETRATE = 100000.0
WINDOWSIZE = 250000.0
USEHAMMING = T
PREEMCOEF = 0.97
NUMCHANS = 24
NUMCEPS = 12
```

また特徴ファイルとしては script.hcopy を使用し内容としては

script.hcopy

```
wave/data0-1.wav mfcc/data0-1.mfc
wave/data0-2.wav mfcc/data0-2.mfc
wave/data0-3.wav mfcc/data0-3.mfc
wave/data1-1.wav mfcc/data1-1.mfc
wave/data1-2.wav mfcc/data1-2.mfc
wave/data1-3.wav mfcc/data1-3.mfc
wave/data2-1.wav mfcc/data2-1.mfc
wave/data2-2.wav mfcc/data2-2.mfc
wave/data2-3.wav mfcc/data2-3.mfc
wave/data3-1.wav mfcc/data3-1.mfc
wave/data3-2.wav mfcc/data3-2.mfc
wave/data3-3.wav mfcc/data3-3.mfc
wave/data4-1.wav mfcc/data4-1.mfc
wave/data4-2.wav mfcc/data4-2.mfc
wave/data4-3.wav mfcc/data4-3.mfc
wave/data5-1.wav mfcc/data5-1.mfc
wave/data5-2.wav mfcc/data5-2.mfc
wave/data5-3.wav mfcc/data5-3.mfc
wave/data6-1.wav mfcc/data6-1.mfc
wave/data6-2.wav mfcc/data6-2.mfc
wave/data6-3.wav mfcc/data6-3.mfc
wave/data7-1.wav mfcc/data7-1.mfc
wave/data7-2.wav mfcc/data7-2.mfc
wave/data7-3.wav mfcc/data7-3.mfc
wave/data8-1.wav mfcc/data8-1.mfc
wave/data8-2.wav mfcc/data8-2.mfc
wave/data8-3.wav mfcc/data8-3.mfc
wave/data9-1.wav mfcc/data9-1.mfc
wave/data9-2.wav mfcc/data9-2.mfc
wave/data9-3.wav mfcc/data9-3.mfc
```

となる。
これは録音した 0 から 9 の wave ファイルを mfcc というフォルダに拡張子を mfc とすることが書かれている。

2.2.4 初期モデルの作成

zero.hmm

```
 o <VecSize> 39 <MFCC_0_D_A>
 h "zero"
<BeginHMM>
<NumStates> 5
<State> 2
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<State> 3
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<State> 4
<Mean> 39
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
<Variance> 39
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
<TransP> 5
0.0 1.0 0.0 0.0 0.0
0.0 0.5 0.5 0.0 0.0
0.0 0.0 0.5 0.5 0.0
0.0 0.0 0.0 0.5 0.5
0.0 0.0 0.0 0.0 0.0
<EndHMM>
```

HMM とは Hidden Markov Model の略であり、確率的非決定性オートマトンである。

また今回は 0 から 9 まで作るので 2 行目の zero の部分をそれぞれの数字に変えて 9 つ同様に作る。

2.2.5 初期値の設定

初期値の設定には HInit というコマンドを用いる。HInit のコマンドは

HInit

HInit [オプション] HMM 名
T 出力トレースレベルを指定
S 学習データリストを指定
M 初期 HMM フォルダ名を指定
H HMM 構成ファイル名を指定
l ラベル名を指定
L ラベルフォルダ名を指定

である。

実際には、HInit -T 1 -S trainlist.txt -M hmm0 -H proto/number.hmm -l number -L label number

number には 0 9 と sil(silent) をそれぞれ入れる。実際に、このコマンドを用いた結果を見てみる。(今回は 6 を例とする。)

また実際には一つにまとめられているが、tex の使用上二つに分けさせていただく。

roku.hmm

o
<STREAMINFO> 1 39
<VECSIZE> 39<NULLD><MFCC_D_A_0><DIAGC>
h "roku"
<BeginHMM>
<NumStates> 5
<State> 2
<Mean> 39
-1.607002e+01 -3.136550e+00 -3.527640e+00 -1.722828e+00 8.386778e-01 -2.640973e-01 2.364814e+00 3.992979e-01 3.254130e+00 2.218182e+00 3.829034e+00 1.110100e+00 5.224752e+01
-9.862473e-03 2.659577e-02 -3.476590e-02 3.656214e-02 -3.275470e-02 -3.411134e-03 8.162678e-02 6.728800e-02 7.522775e-02 -9.567187e-02 -1.181260e-01 1.979477e-02 -1.822452e-02
-4.983091e-03 5.618250e-03 5.753295e-02 6.731976e-02 -1.256003e-02 -4.056766e-03 -8.436010e-03 -1.726922e-02 4.659107e-02 2.300853e-02 -1.621371e-02 -2.650264e-02 -1.566062e-02
<VARIANCE> 39
6.623083e-01 2.275464e+00 2.720420e+00 3.275006e+00 6.796302e+00 1.618636e+01 8.812725e+00 1.173328e+01 9.695393e+00 1.113765e+01 6.760661e+00 1.278183e+01 4.022111e-01
8.428650e-02 3.650138e-01 3.645475e-01 4.913571e-01 9.461271e-01 2.218837e+00 1.171923e+00 1.152024e+00 1.378210e+00 6.822062e-01 4.905273e-01 1.873443e+00 4.205589e-02
1.539480e-02 8.632937e-02 6.678745e-02 9.306833e-02 1.901467e-01 4.795963e-01 2.644285e-01 2.576747e-01 3.052823e-01 1.696136e-01 1.185186e-01 4.219374e-01 1.000000e-02
<GCONST> 5.702334e+01
<STATE> 3
<MEAN> 39
-1.627571e+01 -3.171728e+00 -3.082912e+00 -3.539087e-01 1.232996e+00 1.302350e+00 3.669764e+00 8.909409e-01 3.640772e+00 1.780744e+00 3.553470e+00 9.661123e-01 5.168535e+01
4.858883e-03 -1.096574e-02 1.952368e-02 -1.570482e-02 -7.282193e-02 -7.145162e-02 -1.607958e-02 8.001973e-02 7.932841e-02 8.553989e-02 8.073001e-02 9.498357e-02 1.402565e-03
2.492327e-03 9.438742e-04 -1.103832e-02 -2.316420e-02 -8.760936e-03 -1.268291e-02 8.073286e-03 2.017542e-02 4.550548e-03 -8.212351e-03 -3.958558e-02 -1.686920e-04 3.476190e-03

roku.hmm

<VARIANCE> 39
5.992875e-01 1.566941e+00 2.186272e+00 3.323109e+00 5.759735e+00 6.860231e+00 6.327370e+00 1.451679e+01 1.129030e+01 8.636569e+00 1.146246e+01 7.794979e+00 3.246420e-01
7.639947e-02 1.629448e-01 2.466094e-01 4.452868e-01 6.286964e-01 7.701614e-01 7.185394e-01 2.541940e+00 1.189015e+00 1.282262e+00 1.618738e+00 1.032413e+00 2.541157e-02
1.641713e-02 3.416937e-02 5.335956e-02 9.073702e-02 1.283719e-01 1.442789e-01 1.686950e-01 5.899761e-01 2.441825e-01 2.518194e-01 3.611614e-01 2.025925e-01 1.000000e-02
<GCONST> 5.116494e+01
<STATE> 4
<MEAN> 39
-1.443122e+01 -4.755369e+00 -5.063921e+00 -3.479934e+00 -1.686239e+00 7.367094e-01 5.302240e+00 1.351848e+00 7.343520e+00 4.438652e+00 1.259691e+00 -2.047224e+00 5.368282e+01
-5.530231e-02 -6.349390e-02 -2.135807e-01 -9.653380e-02 1.287086e-01 2.056376e-01 2.934909e-01 1.512398e-01 1.989253e-01 -1.177617e-01 -5.727504e-01 -6.871736e-01 -6.559929e-03 -4.007588e-02 -1.481105e-02 -6.342229e-02 -2.281695e-02 1.274739e-01 1.439306e-01 1.375905e-02 -6.900794e-03 -5.111127e-02 -5.488915e-02 -3.234605e-02 -2.398750e-02 -4.766954e-03 <VARIANCE> 39
5.361225e+00 1.554371e+01 2.265588e+01 2.126947e+01 9.448821e+00 1.411568e+01 1.872291e+01 1.172677e+01 1.562349e+01 6.298062e+00 1.430772e+01 5.742940e+01 8.163157e+00
9.686877e-01 2.279316e+00 3.601607e+00 3.764023e+00 1.352421e+00 1.820420e+00 2.649633e+00 1.113838e+00 2.796088e+00 4.992415e-01 1.805226e+00 8.511679e+00 1.266426e+00
1.810765e-01 4.831476e-01 8.438413e-01 8.791206e-01 3.348599e-01 5.171704e-01 5.698286e-01 2.594590e-01 6.095412e-01 4.911070e-02 4.168569e-01 2.050827e+00 2.663352e-01
<GCONST> 1.032582e+02
<TRANSP> 5
0.000000e+00 1.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
0.000000e+00 9.166667e-01 8.333333e-02 0.000000e+00 0.000000e+00
0.000000e+00 0.000000e+00 9.605263e-01 3.947369e-02 0.000000e+00
0.000000e+00 0.000000e+00 0.000000e+00 7.692308e-01 2.307692e-01
0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
<ENDHMM>

2.2.6 HMM の学習

HRest

HRest [オプション] HMM 名
T 出力トレースレベルを指定
S 学習データリストを指定
M 学習後の HMM フォルダ名を指定
H HMM 構成ファイル名を指定
l ラベル名を指定
L ラベルフォルダ名を指定

Baum-Welth アルゴリズムによってそれぞれの数字におけるお手本となる HMM を作成する。
実際には、HRest -T 1 -S trainlist.txt -M hmm1 -H hmm0/number.hmm -l number -L Label number
number は同様に 0 9 と sil をそれぞれ入れる。これを実際に実行すると、

roku.hmm

o
<STREAMINFO> 1 39
<VECSIZE> 39<NULLD><MFCC_D_A_0><DIAGC>
h "roku"
<BeginHMM>
<NumStates> 5
<State> 2
<Mean> 39
-1.598470e+01 -3.072431e+00 -3.567712e+00 -1.792770e+00 4.949929e-01 7.214302e-02 2.022241e+00 -1.282921e-01 3.170036e+00 1.872616e+00 3.649414e+00 1.069661e+00 5.231638e+01
-3.086665e-02 -8.535717e-02 -1.734930e-01 -3.178370e-02 -1.025233e-02 -9.106028e-02 1.637417e-02 1.488761e-01 -5.168485e-02 -7.694127e-02 -1.494069e-01 -5.138599e-02 -2.662453e-03
-3.063101e-03 -2.478864e-03 5.628845e-02 7.257373e-02 1.467268e-02 1.442834e-02 5.319571e-02 5.948523e-02 7.615205e-02 1.740436e-02 -5.855042e-02 -1.326331e-02 -5.656065e-03
<VARIANCE> 39
5.501991e-01 1.850654e+00 3.369794e+00 3.720270e+00 6.652947e+00 1.573881e+01 6.467906e+00 7.569183e+00 9.070190e+00 9.727133e+00 5.591476e+00 1.217111e+01 3.078845e-01
7.592923e-02 3.308683e-01 3.416337e-01 5.510619e-01 8.026141e-01 1.909318e+00 9.326577e-01 8.841294e-01 1.256679e+00 7.321036e-01 5.834448e-01 1.618331e+00 3.508602e-02
1.337546e-02 6.920389e-02 7.703397e-02 1.242448e-01 1.743087e-01 4.535945e-01 2.375603e-01 1.873879e-01 2.762358e-01 1.936266e-01 1.215561e-01 3.392567e-01 6.560092e-03
<GCONST> 5.325071e+01
<STATE> 3
<MEAN> 39
-1.633046e+01 -3.207698e+00 -3.042130e+00 -2.570768e-01 1.434758e+00 1.189396e+00 3.910338e+00 1.195731e+00 3.702663e+00 1.947492e+00 3.638043e+00 9.816164e-01 5.162401e+01
1.678615e-02 4.757684e-02 9.642603e-02 1.877045e-02 -8.664242e-02 -2.728889e-02 1.476147e-02 3.672279e-02 1.477034e-01 8.331431e-02 1.061444e-01 1.364780e-01 -6.111128e-03
1.783378e-03 5.093783e-03 -1.333395e-02 -2.989902e-02 -2.322941e-02 -2.298809e-02 -2.433563e-02 -1.945203e-02 -1.315277e-02 -6.550127e-03 -1.784349e-02 -6.145502e-03 -1.073165e-03

roku.hmm

<VARIANCE> 39
6.292040e-01 1.758585e+00 1.783903e+00 2.889867e+00 5.535324e+00
7.083543e+00 6.804793e+00 1.634343e+01 1.164622e+01 9.348590e+00
1.231957e+01 7.911707e+00 3.103706e-01
7.982890e-02 1.668465e-01 2.293441e-01 4.112201e-01 6.905931e-01 8.739267e-
01 8.307021e-01 2.741744e+00 1.232272e+00 1.283413e+00 1.608017e+00
1.122688e+00 2.856181e-02
1.755716e-02 4.110297e-02 4.715534e-02 7.295392e-02 1.337130e-01 1.432724e-
01 1.769808e-01 6.404274e-01 2.549593e-01 2.425932e-01 3.696231e-01
2.377525e-01 6.475280e-03
<GCONST> 5.173721e+01
<STATE> 4
<MEAN> 39
-1.443124e+01 -4.755316e+00 -5.063849e+00 -3.479857e+00 -1.686245e+00
7.366293e-01 5.302169e+00 1.351918e+00 7.343489e+00 4.438623e+00
1.259731e+00 -2.047148e+00 5.368278e+01
-5.530998e-02 -6.349972e-02 -2.135759e-01 -9.652862e-02 1.287089e-01
2.056345e-01 2.934792e-01 1.512350e-01 1.989358e-01 -1.177698e-01 -
5.727552e-01 -6.871828e-01 -6.565115e-03
-4.007652e-02 -1.481516e-02 -6.342690e-02 -2.281756e-02 1.274851e-01
1.439439e-01 1.376199e-02 -6.915668e-03 -5.111426e-02 -5.488661e-02 -
3.234636e-02 -2.398589e-02 -4.766238e-03
<VARIANCE> 39
5.361176e+00 1.554361e+01 2.265576e+01 2.126945e+01 9.448665e+00
1.411582e+01 1.872285e+01 1.172683e+01 1.562326e+01 6.298110e+00
1.430756e+01 5.742892e+01 8.163108e+00
9.686726e-01 2.279276e+00 3.601545e+00 3.763967e+00 1.352396e+00
1.820406e+00 2.649596e+00 1.113833e+00 2.796043e+00 4.992468e-01
1.805193e+00 8.511523e+00 1.266404e+00
1.810733e-01 4.831410e-01 8.438270e-01 8.791076e-01 3.348608e-01 5.171707e-
01 5.698189e-01 2.594733e-01 6.095309e-01 4.911028e-02 4.168494e-01
2.050795e+00 2.663305e-01
<GCONST> 1.032579e+02
<TRANSP> 5
0.000000e+00 1.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
0.000000e+00 9.233727e-01 7.662734e-02 0.000000e+00 0.000000e+00
0.000000e+00 0.000000e+00 9.588198e-01 4.118022e-02 0.000000e+00
0.000000e+00 0.000000e+00 0.000000e+00 7.692163e-01 2.307837e-01
0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
<ENDHMM>

となる。

2.2.7 単語認識

単語認識にはHParseとHViteというコマンドを用いる。

HParse

HParse 文法ファイル名　ネットワークファイル名

このコマンドにより number の前後に sil の空間があることを定義する。実際には、
HParse grammar.txt net.slf

HVite

HVite [オプション] 単語辞書ファイル名　HMM リスト名　入力ファイル名
T 出力トレースレベルを指定
H HMM 構成ファイル名を指定
i 認識結果ファイル名を指定
w ネットワークファイル名を指定

HVite -T 1 -S trainlist.txt -H speech/hmmdefs.hmm -i reco.mlf -w net.slf voca.txt hmmlist.txt

2.2.8 認識率の評価

どのくらい認識しているかの評価はHResultsコマンドを用いる。

HResults

HResults [オプション名]　HMM リストファイル名　認識結果ファイル　e ラベル 2 をラベル 1 に置き換える
I 正解ファイル名を指定
L ラベルファイルのフォルダ名を指定

実際には
HResults -T 1 -e "???" sil -I ref.mlf -L label hmmlist.txt reco.mlf > results.txt
とする。

3 まとめ

音声認識の実験を実際に行ってみたが単語認識率は 60 %になってしまった。考えられる原因としては、母数が少ない・ラベル付けがあまいがある。次回取り組

む機会があれば母数を増やしてラベル付けをより丁寧にやって精度を高めたい。また今回は0から9の簡単な数字の認識のみで終わってしまったが、ここから文章や認識するとコンピュータがある動作を行うことができたりするとより良いものとなるだろう。

4　参考文献

荒木雅弘 (2018)「フリーソフトでつくる音声認識システム ー パターン認識・機械学習の初歩から対話システムまで」
htkbook.pdf
http://htk.eng.cam.ac.uk/download.shtml