

TR-I-0366

全音素エルゴディック HMM ユーザーズマニュアル
An All-phoneme Ergodic HMM User's Manual

宮沢康永 嵯峨山茂樹
Yasunaga MIYAZAWA
and Shigeki SAGAYAMA

概要

本レポートでは全音素エルゴディック HMM を用いた教師なし話者適応学習を行なうプログラムについて、その使用方法を説明する。また、C シェルスクリプトで書かれた実際の使用例も添付する。

©ATR 自動翻訳電話研究所
©ATR Interpreting Telephony Research Labs.

目次

| | | |
|-----|-----------------------------|---|
| 1 | はじめに | 1 |
| 1.1 | 使用上の注意 | 1 |
| 2 | プログラムの説明 | 2 |
| 2.1 | 全音素エルゴディック HMM の初期モデルの作成 | 2 |
| 2.2 | 教師なし話者適応学習 | 4 |
| 2.3 | 全音素エルゴディック HMM の話者適応後モデルの分解 | 6 |
| 3 | 実際の例 | 7 |
| 3.1 | インストールとデモ | 7 |

第 1 章

はじめに

全音素エルゴディック HMM を用いて教師なし話者適応学習を行なうプログラムの使用方法について説明する。全音素エルゴディック HMM を用いた教師なし話者適応方式は、教師なし話者適応において確率的に言語情報を用いる方法であり、2、3分程度の発声内容が未知な音声データを用いて高性能な話者適応が可能となる。録音された音声データを認識するようなタスクへの応用が可能である [1][2]。

| | |
|-------------------------------|------------------------|
| Src_HP/Adapt/main | 教師なし話者適応学習プログラム |
| Src/Adapt/make_list | 音声区間切り出しプログラム |
| Src/Bigram/devide_ngram | 音素記号抽出プログラム |
| Src/Make_inithmm/mk_inithmm_m | 全音素エルゴディック HMM 作成プログラム |
| Src/Separate_hmm/sp_inithmm_m | 全音素エルゴディック HMM 分解プログラム |

この章の残りでは CM-HMM ソフトウェアパッケージの使用上の制限について述べる。第 2 章ではデータやファイルのフォーマット、全プログラムの使い方について述べる。第 3 章ではインストールについて述べるとともに、25 単語を用いた話者適応の実行例も示す。

1.1 使用上の注意

このソフトウェアを動作させるためには、DECstation 5000/200(OS は ULTRIX V4.1) 及び HP9000/730 の 2 種類のワークステーションを必要とし、その上での動作確認が行なわれている。

必要とされるメインメモリはタスクに依存する。動作を確認した我々のワークステーションは 64MB(DEC5000/200) 及び 200MB(HP9000/730) のメモリを実装している。

添付してあるデモソフトを実行するには、nawk, uniq, csh などの UNIX プログラムが必要である。

学習のプログラムソースはディレクトリ Src, Src_HP にある。また、実験ツール類はディレクトリ Work, Work_HP にある。

このソフトウェアパッケージで用いる混合連続分布型 HMM の共分散行列は無相関正規分布を仮定した対角成分のみを扱う。

第 2 章

プログラムの説明

2.1 全音素エルゴディック HMM の初期モデルの作成

本章で扱うプログラムは全て DECstation 5000/200 上で動作する。

音素 bigram 確率値を計算するために、テキストファイルから音素記号の種類とその数、連続する音素記号のペアの種類と数を求め、それらの情報を記載した ASCII ファイルを各々作成する。テキストファイルは ASCII ファイルで、単語又は文節を表す音素記号列が記載され、音素記号間にはスペースが、単語間または文節間にはリターンが挿入されているものを用いる。

【テキストファイルの例】

```
.....  
Work/Make_inithmm/List/W5240even.txt  
.....  
a i # 音素名 音素名 ...  
a i ky oou  
a i zy oou  
a i z u  
  (中略)  
w a r i n i  
w a r u  
w a r u k2 u ch2 i  
w a N
```

ディレクトリ Src/Bigram に移りプログラム `devide_ngram` を用いて、テキストファイルから音素記号を単独及びペアで抽出する。更に `sort`, `uniq` を用いて、音素記号の種類と数を記載するファイル及び音素記号ペアの種類と数を記載するファイルを作成する。

【使用例】

1. `devide_ngram` の使用例

```
devide_ngram -n 1 -f TEXTFILE -e out_tmp
```

ここで `-n` この値が 1 の時、音素記号を単独で抽出
2 の時、音素記号をペアで抽出
`-f` テキストファイル名
`-e` 出力ファイル名 (temporary file)

2. 音素記号の種類と数及び音素記号ペアの種類と数を記載するファイル求める例

```
devide_ngram -n 1 -f TEXTFILE -e out_tmp  
sort out_tmp | uniq -c > TEXT.1.dat
```

```
devide_ngram -n 2 -f TEXTFILE -e out_tmp
sort out_tmp | uniq -c > TEXT.2.dat
```

TEXTFILE : テキストファイル名
 TEXT.1.dat: 音素記号の種類と数とを記載したファイル
 TEXT.2.dat: 音素記号ペアの種類と数とを記載したファイル

【音素記号の種類と数を記載したファイルの例】

```
.....
Work/Make_inithmm/List/W5240even.txt.1.dat
.....
126 DQ      # 音素数 音素名
565 H
1714 a
  12 aa
  (中略)
  83 w
  176 y
  243 z
  87 zy
```

【音素記号ペアの種類と数を記載したファイルの例】

```
.....
Work/Make_inithmm/List/W5240even.txt.2.dat
.....
  1 DQ ch2 # 音素ペア数 音素名 音素名
  2 DQ cy
 42 DQ k2
  3 DQ ky
  (中略)
  8 zy o
 39 zy oou
 20 zy u
 14 zy uu
```

このプログラムで用いる音素数は HMM-LR[3] で文節認識を行なうことを前提に 50 個とする。
 このプログラムで用いる音素記号名を表 2.1 に示す。

表 2.1: 音素 HMM の種類

| | |
|------|--|
| 母音 | a i u e o |
| 長母音 | aa ii uu eei oou |
| 子音 | p1 p2 t1 t2 k1 k2 b1 b2 d1 d2 g ng m n N r w y s sh h z ch1 ch2 ts1 ts2 sy hy zy cy py ky by gyl ngy my ny ry (1,2 は語頭、語中を表す) |
| 無音区間 | sil(音声発話区間の前後),DQ(撥音の前部) |

初期モデルとして用いる標準話者の音素の混合連続型 HMM のフォーマットは以下の ATR フォーマットを用いる。

【音素の混合連続型 HMM ファイルの例】

```
.....
inithmm
.....
num_feature= 1
num_omatrix= 3
num_mix= 3          # 混合分布数
feature[1]= 34      # パラメータの次数
mix[0][0]= 3.432870e-01 # observation matrix 0, mixture 0 の分岐確率
  1.470857e+01  6.184570e-01
```

```

1.151042e+00 8.425603e-02
-6.805916e-01 5.766920e-02
1.114223e-01 5.255065e-02
5.922382e-03 4.511616e-02 # observation matrix 0, mixture 0 の34次元分の平均と分散

```

(中略)

```

4
1 0
1 3
6
0 0 8.455312e-01 0 # 状態数
0 1 1.544688e-01 0 # 開始状態数 番号
1 1 8.952486e-01 1 # 終了状態数 番号
1 2 1.047514e-01 1 # アーク数
2 2 8.349881e-01 2 # 開始状態番号 終了状態番号 遷移確率 確率密度関数番号
2 3 1.650119e-01 2 # (以下同様)

```

以上のように作成した2つのファイル、予め学習されている標準話者の上記50音素の混合連続分布型HMMの所在地を記載したファイル、及びプログラムSrc/Make_inithmm/mk_inithmm_mを用いて、全音素エルゴディックHMMの初期モデルを作成する。

【使用例】

```
mk_inithmm_m -p 50 -t TEXT.1.dat -b TEXT.2.dat -h hmmlist > inithmm
```

ここで -p 音素数

-t 音素記号の種類と数とを記載したファイル名

-b 音素記号ペアの種類と数とを記載したファイル名

-h 標準話者HMMリストファイル名

inithmm 出力ファイル名 (全音素エルゴディックHMMの初期モデル)

【標準話者の各音素HMMのリストファイルの例】

```

.....
Work/Make_inithmm/List/MHT_hmm.list
.....
../Hmm/Initial/MHT/hmm_U # 音素HMMファイル名
../Hmm/Initial/MHT/hmm_a
../Hmm/Initial/MHT/hmm_aa
../Hmm/Initial/MHT/hmm_b1
../Hmm/Initial/MHT/hmm_b2
(中略)
../Hmm/Initial/MHT/hmm_z
../Hmm/Initial/MHT/hmm_zy
../Hmm/Initial/MHT/hmm_DQ

```

2.2 教師なし話者適応学習

適応学習はATRの分析ツールで分析された音響パラメータファイルを用いることを前提している。音響パラメータとしては、log power, cepstrum(16次), Δ log power, Δ cepstrum(16次)から成る4種類のファイルを用いている。

各音素ラベルの開始フレーム、継続フレームおよび前後の音素環境を記述したインデックスファイルから、適応学習に用いる各単語又は各文節の音声区間をプログラムSrc/Adapt/make_listにより切り出し、音声区間ファイルに記述する。各単語及び文節が無音区間から始まり無音区間で終ることを前提に適応学習を行うため、音声区間の前後に無音区間を各20フレームずつ付加する。ここで20フレームとは、分析時のフレームシフト長が5msであることから100msに相当する。音声区間ファイルの作成例を示す。

【使用例】

```
make_list -b MAU_B.5mS.idx | head -25 | nawk '{printf"%d %d\n", $3-20, $4+40}' > seqlist
```

ここで -b 区切りを文節ごとまたは単語ごとにするオプション
 MAU_B.5mS.idx インデックスファイル名
 -25 適応学習に用いる単語数、または文節数
 seqlist 出力ファイル (音声区間ファイル)

【音声区間ファイルの例】

```
.....  

Work_HP/Adapt/MAU.B.25.list  

.....  

0 158 # 始端フレーム番号 フレーム長  

158 150  

308 217  

527 170  

698 142  

(中略)  

3815 143  

3959 152  

4113 163  

4277 191
```

上記のように作成された音声区間ファイル、音響パラメータファイル、及び前章のように作成された全音素エルゴディック HMM の初期モデルを用いて、教師なし適応学習を行なう。以下の話者適応プログラムは HP9000/730 上で動作する。

【使用例】

```
Src_HP/Adapt/main \  

-i inithmm \  

-o outhmm \  

-x MAU.B.25.list \  

-l log.tbl.NEW \  

-p Coninfo \  

-F 1.4 \  

-m 10 \  

-f MAU_B.5mS.lpow 1 MAU_B.5mS.cep16 16 \  

  MAU_B.5mS.dpow 1 MAU_B.5mS.dcep16 16
```

ここで -i 初期モデルのファイル名
 -o 出力ファイル名の一部
 出力ファイル名は [_itr 繰り返し回数] が付加される
 出力ファイル名の例: outhmm_itr1 (繰り返し回数=1)
 -x 音声区間ファイル
 -l 対数テーブル
 -p パラメータの重み係数ファイル
 -F ファジネス
 -m 学習回数 (デフォルト: 10回)
 -f 入力音響パラメータファイル名 入力パラメータ次元数
 (ここでは4つのファイルで構成されている)

【音声区間ファイルの例】

```

.....
Work_HP/Adapt/List/Coninfo
.....
.lpow  lpow  1   0.02 # 対数パワー file の拡張子  dummy  次数  重み係数
.cep16 cep   16  6.0  # ケプストラム file の拡張子  dummy  次数  重み係数
.dpow  dpow  1   0.1  # Δ対数 file の拡張子    dummy  次数  重み係数
.dcep16 dcep  16  5.0  # Δケプストラム file の拡張子 dummy  次数  重み係数

```

2.3 全音素エルゴディック HMM の話者適応後モデルの分解

前章の学習プログラムで話者適応された全音素エルゴディック HMM を各音素 HMM に分解する。

【使用例】

```

Src/Separate_hmm/sp_inithmm_m  \
    -p 50                       \
    -i outhmm_itr10            \
    -h hmm.list                 \
    -o out_hmm.list

```

ここで -p 音素数

- i 話者適応後の全音素エルゴディック HMM のファイル名
- h 話者適応前の標準話者の各音素 HMM のリストファイル
- o 話者適応後の各音素 HMM のリストファイル (出力ファイルリスト)

【話者適応後の各音素 HMM のリストファイルの例】

```

.....
Work/Separate_hmm/List/out_MHT2MAUB25_f1.4_itr8.list
.....
# 音素 HMM ファイル名
.../Hmm/Adapt/MHT2MAUB25/iter_8/hmm_B
.../Hmm/Adapt/MHT2MAUB25/iter_8/hmm_a
.../Hmm/Adapt/MHT2MAUB25/iter_8/hmm_aa
.../Hmm/Adapt/MHT2MAUB25/iter_8/hmm_b1
.../Hmm/Adapt/MHT2MAUB25/iter_8/hmm_b2
(中略)
.../Hmm/Adapt/MHT2MAUB25/iter_8/hmm_z
.../Hmm/Adapt/MHT2MAUB25/iter_8/hmm_zy
.../Hmm/Adapt/MHT2MAUB25/iter_8/hmm_DQ

```

第 3 章

実際の例

3.1 インストールとデモ

実行するにはディスクにおよそ 16MB の空きスペースが必要です。
DECstation 5000/200 上で以下のコマンドをタイプしてください。

```
$ mkdir ADAP-TRAIN          # create directory ADAP-TRAIN
$ cd ADAP-TRAIN             # go to newly created directory
$ tar xvf                   # get package from the default cartridge drive
$ set THISDIR = 'pwd'
$ cd $THISDIR/Src/Bigram    #
$ make -f devide_ngram.make # create devide_ngram
$ cd $THISDIR/Src/Make_inithmm #
$ make -f mk_inithmm_m.make # create mk_inithmm_m
$ cd $THISDIR/Src/Separate_hmm #
$ make -f sp_inithmm_m.make # create sp_inithmm_m
$ cd $THISDIR/Src/Adapt    #
$ make -f makefile         # create make_list
```

次に HP9000/730 上で以下のコマンドをタイプしてください。

```
$ mkdir ADAP-TRAIN          # create directory ADAP-TRAIN
$ cd ADAP-TRAIN             # go to newly created directory
$ tar xvf                   # get package from the default cartridge drive
$ cd Src_HP/Adapt           #
$ make -f Makefile          # create main
```

次のコマンドでデモプログラムが実行できます。

まず、DECstation 5000/200 上で全音素エルゴディック HMM の初期モデルを作成する。

```
$ cd Work/Make_inithmm     # go to make inithmm directory
$ mk_inithmm.csh           # make inithmm
```

DECstation 5000/200 上に作成された全音素エルゴディック HMM の初期モデル

ADAP-TRAIN/Hmm/Initial/MHT/inithmm_MHT_even を

HP9000/730 上のディレクトリ

ADAP-TRAIN/Hmm/Initial/MHT

にコピーする。次に、HP9000/730 上で話者適応学習を行なう。

```
$ cd Work_HP/Adapt          # go to adaptation directory
$ AdaptTrain.csh           # run adaptation
```

このプログラムは HP9000/730 上で、約2時間かかります。HP9000/730 上に作成された全音素エルゴディック HMM の話者適応後のファイル

```
ADAP-TRAIN/Hmm/Adapt/MHT2MAUB25/hmmMHT2MAUB25_f1.4_itr8 を
```

```
DECstation 5000/200 上のディレクトリ
```

```
ADAP-TRAIN/Hmm/Adapt/MHT2MAUB25
```

```
にコピーする。次に、DECstation 5000/200 上で全音素エルゴディック HMM を分解する。
```

```
$ cd Work/Separate_hmm      # go to separate hmm directory
$ separate_m.csh           # separate hmm
```

以上により、DECstation 5000/200 上のディレクトリ

```
Hmm/Adapt/MHT2MAUB25/iter_8
```

に話者適応後の各音素 HMM が作成されます。このデモの話者適応学習では、ファジネスの値は 1.4、学習の繰り返し数は 8 とした。

参考文献

- [1] 宮沢 康永, 大倉 計美, 嵯峨山 茂樹: “全音素エルゴディック HMM による教師なし話者適応,” 信学技報, SP92-75, pp. 15-20 (1992.10).
- [2] 宮沢 康永, 嵯峨山 茂樹: “全音素エルゴディック HMM を用いた教師なし話者適応における初期モデルの検討,” 音講論集, 2-4-10, pp. 37-38 (1993.3).
- [3] 山口耕市, 嵯峨山茂樹: “混合連続分布型 HMM を用いた HMM-LR 連続音声認識”, 音講論集, 1-P-5, pp.113-114 (1992.3).