

TR-I-0364

混合連続分布型 HMM-LR ユーザーズマニュアル
Continuous Mixture HMM-LR User's Manual

山口 耕市
Kouichi YAMAGUCHI

1993 年 3 月

概要

本レポートでは、混合連続分布型 HMM-LR 連続音声認識システムにおける音素モデルの学習および連続音声の認識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	HMM 音素モデルの作成	2
2.1	音素ラベルファイルの作成	2
2.2	HMM の学習	3
2.3	継続時間長ファイルの作成	4
3	HMM-LR の実行	5
3.1	引用ファイルの整備	5
3.1.1	HMM 音素モデルリストの作成	5
3.1.2	音声データの指定	6
3.2	プログラムの実行	7
3.2.1	オプションの指定	7
3.2.2	実行例	8

第 1 章

はじめに

混合連続分布型 HMM-LR の学習・認識を行なうプログラムの使用方法について説明する。混合連続分布型 HMM-LR は離分布散型 HMM-LR に比べ、計算量が多いものの高精度な音素モデルを構築することができる [1, 4, 6]。また、話者適応や不特定話者への発展性も期待されている。

このソフトウェアパッケージは C のプログラムと C シェルスクリプトからできている。この章の残りでは CMHMM-LR 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 章では音素モデル作成のためのラベルファイルのフォーマット変換と学習の仕方について述べる。第 3 章では HMM-LR 認識プログラムの使い方について述べるとともに、文節音声認識実験の実行例も示す。

1.1 使用上の注意

このソフトウェアは DECstation 5000/25(OS は ULTRIX V4.1) 上でコンパイルされ、動作が確認されている。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 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, sed, csh などの UNIX プログラムが必要である。音素モデルの作成用プログラム、ツール類はディレクトリ Chmm にある。また、HMM-LR 用実験プログラム、ツール類はディレクトリ CMHMM-LR にある。

なお、このソフトウェアパッケージでは連続 HMM は無相関正規分布を仮定した対角成分のもののみを扱っている。

第 2 章

HMM 音素モデルの作成

ここで扱うラベルファイルおよび HMM については、文献 [5] の“CM-HMM(混合連続分布型 HMM) ユーザーズマニュアル” (ATR Technical Document TR-I-0277) と同じ仕様に基づいているので、そちらを参照してください。

2.1 音素ラベルファイルの作成

ディレクトリ Chmm/label の下で以下のツールを使う。

lab_convY	ラベル変換プログラム
gen_yama.csh	ラベル変換サブシェルスクリプト
gen_yama_all.csh	ラベル変換シェルスクリプト
phon_all.csh	音素ラベル抽出シェルスクリプト

まず、gen_yama_all.csh を用いて標準的なインデックスファイル XXX_*.5mS.idx から HMM-LR で扱っている音素ラベルファイル XXX_*.5mS.yama に変換する。ここで、“XXX” は話者名、“*” はデータセット名に相当する。

【使用例】

```
gen_yama_all.csh MAU
```

ここで MAU は話者名

【インデックスファイルの例】

```
Chmm/label/MAU/MAU_1.5mS.idx  
Chmm/label/MAU/MAU_B.5mS.idx  
Chmm/label/MAU/MAU_DSB.5mS.idx
```

【変換後の音素ラベルファイルの例】

```
Chmm/label/MAU/MAU_1.5mS.yama  
Chmm/label/MAU/MAU_B.5mS.yama  
Chmm/label/MAU/MAU_DSB.5mS.yama
```

次に、phon_all.csh を用いて音素ラベルファイル XXX_*.5mS.yama から音素ラベルを抽出する。重要語 5240 単語、音素バランス 216 単語および複合語を許さない文節発声データ (SB1,2,3,4) の 3 種類の音素ラベルファイルが生成される。ディレクトリ Chmm/label/話者名 の下に、音素ラベルファイル毎に 1 音素につき 1 ファイル生成される。

ただし、文節認識の対象を SB3 とするときにはオープン実験にするために、文節発声データのインデックスファイル (例: MAU_DSB.5mS.idx) から SB3 の部分をエディタ等で削除しておく必要がある。

【使用例】

```
phon_all.csh MAU
```

ここで MAU は話者名

2.2 HMM の学習

ディレクトリ Chmm/Work で学習用シェルスクリプト train_misc.csh を用いる。HMM はディレクトリ Chmm/Hmm/ 話者名 の下に、混合数毎に 1 音素につき 1 ファイル生成される。

【使用例】

```
train_misc.csh MAU
```

ここで MAU は話者名

このシェルスクリプトには以下のツールが含まれている。

Chmm/Tool/make_inithmm	初期化プログラム
Chmm/Src/Train/main	学習プログラム
Chmm/Work/trainV+B_3.csh	母音学習サブシェルスクリプト
Chmm/Work/trainC+B.csh	子音学習サブシェルスクリプト
Chmm/Work/trainQ+B.csh	無音学習サブシェルスクリプト

なお、この例では母音モデルとして子音と同様の 4 状態 3 ループの HMM を採用している。

上の例では重要語 5240 単語と音素バランス 216 単語を学習対象データセットとしている。より高い文節音声認識を望む場合は、これらのデータセットに複合語を許さない文節発声データセット (SB1,2,4) を追加することができる。このとき、シェルスクリプトとしては以下のツールを用いる。

Chmm/Tool/make_inithmm	初期化プログラム
Chmm/Src/Train3/main	学習プログラム
Chmm/Work/trainV++_3.csh	母音学習サブシェルスクリプト
Chmm/Work/trainC++_3.csh	子音学習サブシェルスクリプト

次に、無声化母音モデルを作成する。これは通常之母音モデルにヌル遷移を追加したもので、ディレクトリ Chmm/Hmm/ 話者名 /M 混合数 で生成用シェルスクリプト genS3U.csh を用いる。

【使用例】

```
cd MAU/M12S3B
../genS3U.csh hmm_i hmm_Ui
../genS3U.csh hmm_u hmm_Uu
```

ここで MAU は話者名、M12S3B は混合数のディレクトリ名

最後に、長母音モデルを作成する。通常に学習された長母音モデルをそのまま用いてもよいが、ここでは通常学習された短母音モデルの確率密度関数に通常学習された長母音モデルの状態遷移確率を組み合わせたものを用いる。長母音の学習サンプルが少ないためこのような措置がとられた。特に、/a:/ と /i:/ には有効である。ディレクトリ Chmm/Hmm/ 話者名 /M 混合数 で生成用シェルスクリプト genlongV.csh を用いる。

【使用例】

```
cd MAU/M12S3B
../../genlongV.csh
```

ここで MAU は話者名、M12S3B は混合数のディレクトリ名

2.3 継続時間長ファイルの作成

継続時間長ファイルは、HMM-LR で音素照合区間制御および HMM の状態毎の継続時間長制御に用いられる。

ディレクトリ Chmm/Dur/Exe で学習用シェルスクリプト cal_misc.csh を用いる。継続時間長ファイルはディレクトリ Chmm/Dur/ 話者名 の下に、混合数毎に 1 音素につき複数ファイル生成される。

【使用例】

```
cal_misc.csh MAU
```

ここで MAU は話者名

このシェルスクリプトには以下のツールが含まれている。

Chmm/Dur/Tool/make_data	データ作成プログラム
Chmm/Src/main	継続時間算出プログラム
Chmm/Dur/Exe/cal_durV+B.csh	母音用シェルスクリプト
Chmm/Dur/Exe/cal_durC+B.csh	子音用シェルスクリプト
Chmm/Dur/Exe/cal_durQ+B.csh	無音用シェルスクリプト
Chmm/Dur/Exe/dchangeV_cdsb.csh	母音用変換シェルスクリプト
Chmm/Dur/Exe/dchangeC_cdsb.csh	子音用変換シェルスクリプト
Chmm/Dur/Exe/dchangeQ_cdsb.csh	無音用変換シェルスクリプト
Chmm/Dur/Exe/cp_dur_S3U.csh	無声化母音用サブシェルスクリプト

第 3 章

HMM-LR の実行

ここで扱う HMM-LR 連続音声認識システムの技術的内容については、文献 [2] を参照してください。プログラムの取り扱い説明については文献 [3] の離散分布型 HMM-LR と同じ仕様に基づいている。よって本レポートでは、文献 [3] との相違点を中心に述べる。

なお、この章では固定ビームサーチ版について扱う。

3.1 引用ファイルの整備

3.1.1 HMM 音素モデルリストの作成

前章で作成した HMM 音素モデルファイルおよび継続時間長ファイルのリストをエディタ等で作成する。

これは HMM-LR の `-m` オプションである。

【音素モデルリストの例】

```
.....  
CMHMM-LR/HMM/MAU/model_list  
.....  
Chmm/Hmm/MAU/M14B/hmm_S s 1  
Chmm/Dur/MAU/M14B/dur.cdsb_S  
Chmm/Hmm/MAU/M14B/hmm_SH sh 1  
Chmm/Dur/MAU/M14B/dur.cdsb_SH  
Chmm/Hmm/MAU/M14B/hmm_H h 1  
Chmm/Dur/MAU/M14B/dur.cdsb_H  
Chmm/Hmm/MAU/M14B/hmm_z z 1  
Chmm/Dur/MAU/M14B/dur.cdsb_z  
Chmm/Hmm/MAU/M3B/hmm_CH1 ch1 1  
Chmm/Dur/MAU/M3B/dur.cdsb_CH1  
Chmm/Hmm/MAU/M7B/hmm_CH2 ch 1  
Chmm/Dur/MAU/M7B/dur.cdsb_CH2  
Chmm/Hmm/MAU/M7B/hmm_TS1 ts1 1  
Chmm/Dur/MAU/M7B/dur.cdsb_TS1  
Chmm/Hmm/MAU/M14B/hmm_TS2 ts 1  
Chmm/Dur/MAU/M14B/dur.cdsb_TS2  
Chmm/Hmm/MAU/M3B/hmm_P1 p1 1  
Chmm/Dur/MAU/M3B/dur.cdsb_P1  
Chmm/Hmm/MAU/M3B/hmm_P2 p 1  
Chmm/Dur/MAU/M3B/dur.cdsb_P2  
Chmm/Hmm/MAU/M14B/hmm_t1 t1 1  
Chmm/Dur/MAU/M14B/dur.cdsb_t1  
(中略)
```

```

Chmm/Hmm/MAU/M3B/hmm_gy1 gy 1
Chmm/Dur/MAU/M3B/dur.cdsb_gy1
Chmm/Hmm/MAU/M3B/hmm_ngy ngy 1
Chmm/Dur/MAU/M3B/dur.cdsb_ngy
Chmm/Hmm/MAU/M2B/hmm_my my 1
Chmm/Dur/MAU/M2B/dur.cdsb_my
Chmm/Hmm/MAU/M3B/hmm_ny ny 1
Chmm/Dur/MAU/M3B/dur.cdsb_ny
Chmm/Hmm/MAU/M3B/hmm_ry ry 1
Chmm/Dur/MAU/M3B/dur.cdsb_ry
Chmm/Hmm/MAU/M7S3B/hmm_DQ Q 1
Chmm/Dur/MAU/dur15/dur.cdsb_-
Chmm/Hmm/MAU/M15S3B/hmm_Q1 Q1 1
Chmm/Dur/MAU/dur15/dur_Q1
Chmm/Hmm/MAU/M15S3B/hmm_Q2 Q2 1
Chmm/Dur/MAU/dur15/dur_Q1

```

3.1.2 音声データの指定

文献[3]における音声データの指定は、ここでは2つのファイル（発声リストと音響パラメータリスト）によって行なう。

発声リストの作成

まず、`make_list` を用いて標準的なインデックスファイル `XXX*.5mS.idx` から HMM-LR で扱うことのできる発声リストのファイル形式に変換する。これは HMM-LR の引数なしオプションである。

以下の使用例では文節認識の対象を SB3 にするために、文節発声データのインデックスファイル `MAU_DSB.5mS.idx` から SB3 の部分のみをエディタ等で抽出したファイル `MAU_SB3.5mS.idx` を用いている。

【使用例】

```

cd CMHMM-LR/DATA
../Tool/make_list -b MAU_SB3.5mS.idx > MAU_List_SB3

```

ここで MAU は話者名、SB3 はデータセット名

【発声リストの例】

```

::::::::::::::::::
CMHMM-LR/DATA/MAU_List_SB3
::::::::::::::::::
001          daiikkai    96252  169
002          tsuuyaku    96591  127
003          deNwa      96809  104
004          kokusai    97056  110
005          kaigini     97268   93
006          saNkano     97531  119
007          tourokuo    97755  126
008          gokibousareru 98003  202
009          katawa     98321   73
010          shoteino    98669  133

```


(中略)

270	shippitsushite	158250	189
271	itadakimasu	158500	143
272	soudesuka	158685	126
273	wakarimashita	158948	133
274	dewa	159272	74
275	moushikomi	159441	134
276	youshino	159646	112
277	keN	159832	52
278	yoroshiku	160036	107
279	onagai	160228	94
280	moushi	160414	97
281	agemasu	160592	123

音響パラメータリストの作成

認識対象音声データの音響パラメータファイルのリストをエディタ等で作成する。この例では4つのファイルからなる34次元パラメータを用いている。

これはHMM-LRの-Dオプションで指定する。

【音響パラメータリストの例】

```

.....
CMHMM-LR/DATA/para_list.MAU
.....
data/DSB/MAU/MAU_DSB.5mS.lpow
data/DSB/MAU/MAU_DSB.5mS.cep16
data/DSB/MAU/MAU_DSB.5mS.dpow
data/DSB/MAU/MAU_DSB.5mS.dcep16

```

3.2 プログラムの実行

3.2.1 オプションの指定

文献 [3] の離散分布型 HMM-LR に追加または改定されたものについて説明する。

分散の拡大 (-v オプション)

HMM の確率密度関数の分散に対する定数倍操作である。これにより、発話様式の違いに対するロバスト性を向上させることができる。通常、1.0 から 2.0 までの値を採用する。

分散のフロアリング (-f オプション)

HMM の確率密度関数の分散に対するフロアリング操作である。これにより、極端に小さい分散による尤度の変動を防ぐ。default では各成分毎にフロアリング値が与えられる。

出力確率のフロアリング (-p オプション)

HMM の確率密度関数の出力値に対するフロアリング操作である。実験によれば、この操作の効果はあまりなかったので通常は使用しない。つまり、0 を与えておく。

前後の無音区間 (-A オプション)

音声区間の前後に何フレームの無音区間を含めるかを指定する。離散分布型 HMM-LR では単位は msec であった。

3.2.2 実行例

CMHMM-LR/Src/fixbeam/main を用いる。下の例では、結果をファイル result に書き込んでいる。HMM の尤度計算に要した時間も表示している。誤りサンプルや認識率は CMHMM-LR/Tool/score で、処理時間や音素照合回数の統計量等は CMHMM-LR/Tool/count_stat.awk でそれぞれ見ることができる。なおこの例では、継続時間制御は音素照合区間の決定に対してのみ行ない、HMM の状態毎の継続時間長制御は行なっていない。

【使用例】

```
cd CMHMM-LR/Exe
../Src/fixbeam/main \
../DATA/MAU_List_SB3 \
-g ../GRAMMAR/sp2 \
-m ../HMM/MAU/model_list \
-D ../DATA/para_list.MAU \
-P 5 -M 1 -B 600 -b 64 -c 3000 \
-s 2 -S 3 -W 0 -A 7 -v 1.2 -f 0 -p 0 \
| tee result
```

【結果の例】

```
.....:
CMHMM-LR/Exe/result
.....:
(001) 150549 213 |daiikkai|
Making likelihood map: CPU-time = 21748 msec, Elapsed-time = 22 sec.
*****
Parsing time: CPU-time = 44493 msec, Elapsed-time = 45 sec.
Total-verify = 11923 Depth = 23
 1: daiiQ-kai (prob = -48.17049)
 2: daitai (prob = -47.88851)
 3: daiiQ-kai-ni (prob = -47.79761)
 4: daiiQ-kai-o (prob = -47.78411)
 5: daiiQ-kai-e (prob = -47.78381)
*****

(002) 151018 184 |tsuuyaku|
Making likelihood map: CPU-time = 18655 msec, Elapsed-time = 19 sec.
*****
Parsing time: CPU-time = 38657 msec, Elapsed-time = 38 sec.
Total-verify = 11335 Depth = 20
 1: tsuuyaku (prob = -44.89896)
 2: tsuuyaku-e (prob = -44.62077)
 3: tsuuyaku-o (prob = -44.53977)
 4: tsuuyaku-de (prob = -44.38088)
 5: tsuuyaku-ne (prob = -44.35348)
*****

(003) 151394 132 |deNwa|
Making likelihood map: CPU-time = 13417 msec, Elapsed-time = 13 sec.
*****
Parsing time: CPU-time = 27759 msec, Elapsed-time = 28 sec.
```

```
Total-verify = 10800 Depth = 15
  1: deNwa (prob = -46.64427)
  2: deNwa-ga (prob = -46.27629)
  3: deNwa-o (prob = -46.21939)
  4: deNwa-wa (prob = -46.21479)
  5: zeN-wa (prob = -46.19089)
*****
```

(以下省略)

【認識率の例】

```
../Tool/score result
Missed 164. chikatetsude
Missed 183. chikatetsuto
Missed 212. zikokuhyouga
***** Score of LR-HMM *****
```

```
Total phrases: 278
Phrases parsed correctly: 275
```

```
1: 255 (91.73%) / 255 (91.73%)
2: 14 (5.04%) / 269 (96.76%)
3: 4 (1.44%) / 273 (98.20%)
4: 1 (0.36%) / 274 (98.56%)
5: 1 (0.36%) / 275 (98.92%)
```

```
nawk -f ../Tool/count_stat.awk result
Number of verify
Total phrases: 278
Total verify: 3161974
Average verify: 11374.0
Parsing
Total Elapsed-time: 9692 sec
Average Elapsed-time: 34.9 sec
Making likelihood map
Total Elapsed-time: 4792 sec
Average Elapsed-time: 17.2 sec
```

参考文献

- [1] 花沢利行、川端豪、鹿野清宏：“HMMによる音韻認識実験の結果”，ATR Technical Report, TR-I-0147 (1990.2).
- [2] 北研二、川端豪、斉藤博昭：“HMM音韻認識と拡張LR構文解析法を用いた連続音声認識”，情報処理学会論文誌, Vol.31, No.3, pp.472-479 (1990.3).
- [3] 北研二：“HMM-LR ユーザーズ・マニュアル”，ATR Technical Report, TR-I-0246 (1992.3).
- [4] 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).
- [5] 山口耕市、Harald Singer、嵯峨山茂樹：“CM-HMM(混合連続分布型HMM)ユーザーズマニュアル”，ATR Technical Document, TR-I-0277 (1992).
- [6] K. Yamaguchi, K. Kita, S. Sagayama and F. K. Soong: “Continuous Mixture HMM-LR Using the A* Algorithm for Continuous Speech Recognition”, *Proc. ICSLP-92*, We.sAM.1.2, pp.301-304 (1992.10).