

TR-IT-0100

CCLによる不特定話者混合連続分布型 HMnet 作成ユーザ
ズマニュアル Ver1.0

Continuous Mixture HMnet for Speaker-
Independent Speech Recogni-
tion by CCL User's Manual

小坂 哲夫

Tetsuo KOSAKA

概要

本レポートでは CCL(HMMs' Composition and Clustering)による不特定話者用混合連続分布型 HMnet(Hidden Markov Network)の音素モデル作成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。本レポートで扱う範囲は音素モデルの作成のみで音声パラメータファイルの作成部及び、認識システムは含まない。

©ATR 音声翻訳通信研究所

©ATR Interpreting Telecommunications Res. Labs.

目次

1	はじめに	1
1.1	ディレクトリーの構成	1
1.2	動作に必要なファイル等	1
1.3	使用上の注意	2
2	プログラム及びファイルの説明	3
2.1	不特定話者用 CM-HMnet	3
2.2	初期 HMnet	3
2.3	音響パラメータファイルと音素ラベルファイル	3
2.4	不特定話者用 CM-HMnet の作成法	4
2.4.1	概要	4
2.4.2	シェルプログラムの使用法	4
2.5	HMnet のクラスタリング	5
2.5.1	概要	5
2.5.2	シェルプログラムの使用法	5
2.5.3	belong.csh	6

第 1 章

はじめに

CCL による混合連続分布型 HMnet (CM-HMnet) の音素モデルの作成を行なうプログラムの使用方法について説明する。HMnet は状態を音素間で共有することを特徴とした音素環境依存モデルであり [1]、本パッケージは CCL 法を用いて不特定話者用 HMnet を作成するものである。またこれに関連して本プログラムには HMnet のクラスタリングを行なうプログラムも含まれている。本プログラムの動作には別に SSS プログラム及び分析プログラムが必要である。また SSS-LR プログラムを購入した場合、不特定話者認識システムとして使えるよう、自由発話音声で学習した不特定話者音素モデルも含まれている。CCL 法については、文献 [4] を参考のこと。

1.1 ディレクトリーの構成

```
/data/atrh23/CCL
|
+- CCLexe CCL 実行ファイル格納ディレクトリー
+- Model HMnet モデルファイル
| |
| +- FTK 女性話者 FTK 用 HMnet
| +- MHT 男性話者 MHT 用 HMnet
|
+- List クラスタリング結果ファイル
|
+- Clustexe クラスタリング実行ファイル格納ディレクトリー
```

1.2 動作に必要なファイル等

本プログラムの実行には以下のファイルやプログラムが必要です。

/home/atrh23/kosaka/ARCADIA/BIN.HP-UX/expand_sample 音声入力ファイル整形プログラム

/home/atrh23/kosaka/tmp/SSS/BIN.HP-UX/Exe.adapt_HMnet HMnet 用話者適応プログラム

/home/atrh23/kosaka/tmp/SSS/BIN.HP-UX/Exe.compose_HMnet HMnet 用合成プログラム

/home/atrh23/kosaka/tmp/SSS/BIN.HP-UX/Exe.delete_HMnet HMnet 用音素消去プログラム

/home/atrh16/singer/SRCONV/BIN.HP-UX/srconv ダウンサンプリングプログラム

/home/atrh16/singer/MIKOSHI/BIN.HP-UX/Wave2Para 分析プログラム

1.3 使用上の注意

このソフトウェアは HP9000/735(OS は HP-UX 9.05) 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学習データ量に依存する。

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

第 2 章

プログラム及びファイルの説明

2.1 不特定話者用 CM-HMnet

本パッケージには、別売の SSS-LR プログラムと組み合わせて、不特定話者音声認識のシステムが組めるよう、不特定話者用の自由音声データで学習された音素モデル (CM-HMnet) が含まれている。音素モデルはアスキーテキストであり more などでも確認できる。

```
Model/MULTI_without_FYOMA      # 男女話者用、状態数 201
```

SSS-LR を不特定話者で動かす場合は SSS-LR のマニュアルを参照のこと。

2.2 初期 HMnet

本パッケージには CCL 用に作成済の初期 HMnet が以下のように含まれている。

```
Model/FTK/HMnet_filled.200 # 状態数 200、混合数 1 の話者 FTK 用 HMnet (無音は含  
まず)  
Model/FTK/HMnet_filled.300 # 状態数 300、混合数 1 の話者 FTK 用 HMnet (無音は含  
まず)  
Model/FTK/HMnet_filled.400 # 状態数 400、混合数 1 の話者 FTK 用 HMnet (無音は含  
まず)  
Model/FTK/HMnet_filled.600 # 状態数 600、混合数 1 の話者 FTK 用 HMnet (無音は含  
まず)  
Model/MHT/HMnet_filled.200 # 状態数 200、混合数 1 の話者 MHT 用 HMnet (無音は含  
まず)  
Model/MHT/HMnet_filled.300 # 状態数 300、混合数 1 の話者 MHT 用 HMnet (無音は含  
まず)  
Model/MHT/HMnet_filled.400 # 状態数 400、混合数 1 の話者 MHT 用 HMnet (無音は含  
まず)  
Model/MHT/HMnet_filled.600 # 状態数 600、混合数 1 の話者 MHT 用 HMnet (無音は含  
まず)  
Model/HMM.-.10mix          # 状態数 1、混合数 10 の無音用 HMnet
```

2.3 音響パラメータファイルと音素ラベルファイル

音響パラメータと音素ラベルはそれぞれ別のファイルとなっている。音響パラメータとしては、図 2.1 に示すように、通常、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。ラベルファイルは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ものである。

log power, cepstrum(16次), Δ log power, Δ cepstrum(16次) から成る34次元ベクトルの作成方法については別売の分析ツールを参照のこと。

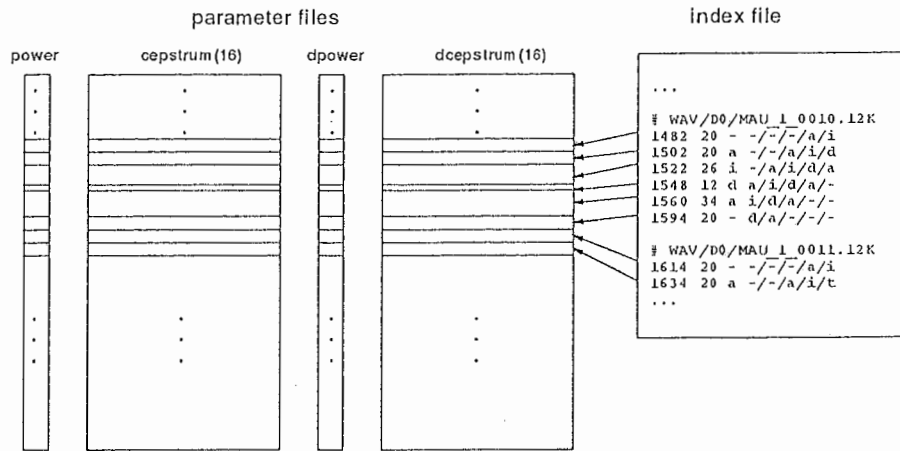


図 2.1: 音響パラメータとラベル (インデックス) ファイルの関係

2.4 不特定話者用 CM-HMnet の作成法

2.4.1 概要

CCL 法では特定話者モデルを合成して任意の混合数の不特定話者モデルを作成する。詳しくは文献 [4] を参照のこと。以下に概要を示す。

ポーズモデルの合成 初期モデルにポーズモデルを付加する。

特定話者モデルの作成 初期モデルに対し VFS 法を利用して複数の話者の比較的少量のデータによりそれぞれ適応をおこない、複数話者分の HMnet を作成する。

ポーズモデルの削除 この複数話者分の HMnet のそれぞれに対しポーズモデルを削除する。

CCL 法による不特定話者モデルの作成 CCL 法を用いて特定話者のモデルを合成して、不特定話者のモデルを作成する。

ポーズモデルの合成 再びポーズモデルを付加する。

2.4.2 シェルプログラムの使用法

CCLexex/gousei2.csh

使用目的 本プログラムではまず入力モデルを話者適応により適応して、複数人分の特定話者モデルを作成する。このとき使用するデータは C セット SG の A (50 文)。次にこれを合成して、不特定話者モデルを作成する。さらに 10 混合の無音モデルを付加する。

使用法 gousei2.csh クラスタリングリスト 入力モデル 出力モデル 状態数

ここでクラスタリングリストは、合成のために話者をクラスタリングしたリストである。フォーマットは以下の通りである。

クラスタ数

クラスタ 1 の話者 1 クラスタ 1 の話者 2 クラスタ 1 の話者 3
 クラスタ 2 の話者 1 クラスタ 2 の話者 2 クラスタ 2 の話者 3
 クラスタ 3 の話者 1 クラスタ 3 の話者 2 クラスタ 3 の話者 3

入力モデルは男性用及び女性用の初期 HMnet をそれぞれ拡張子 .male または .female をつけて用意する。

出力モデルは合成後の不特定話者モデル名を書く。

状態数には無音モデルを除いた状態数を設定する。

使用例 Exe/test.csh に使用例があるので参照のこと。

CCLexe/test.csh

使用目的 gousei2.csh のテストプログラム。まずデータベース上の C セットデータを各自のホームディレクトリーの /Cset にリンクする。その後 15 人を 5 クラスにクラスタリングした結果を用いて、5 混合 200 状態+無音 1 状態の HMnet を作成する。

使用法 test.csh と入力するのみ。合成したモデルは../Model/newmodel.200 として作成される。

2.5 HMnet のクラスタリング

2.5.1 概要

モデル合成のためには、HMnet のクラスタリング結果が必要である。本節では HMnet のクラスタリング法について説明する。ここで用いた HMnet のクラスタリングアルゴリズムは、歪みが最大となるクラスタを順次分割する方法をとった。クラスタリング対象の HMnet が同一の構造を持つことを仮定する。HMnet の初期状態確率や状態遷移確率を無視し、出力確率のみで距離を定義する。出力確率間の距離は Bhattacharyya 距離を用いた。詳細は文献 [4] を参考のこと。

2.5.2 シェルプログラムの使用法

clust/clust-sil

使用目的 HMnet を Bhattacharyya 距離に従ってクラスタリングするプログラム。特徴として、(1)HMnet 間のアライメントはとらず、同一トポロジーにある出力分布間の距離を使用。(2)クラスタリングアルゴリズムは SPLIT 法のものを利用。clust-sil.csh は 10 混合、1 状態の無音モデルを付加したものに对应。(距離計算のとき第 1 状態を無視)

使用法 clust condition リスト スレシヨルド

スレシヨルドの大小によって、クラスタリング数が増減する。

コンデションリストのフォーマットは以下の通り。

モデル総数
 モデル名1
 モデル名2

使用例 `clust-sil ../List/condition.285 160.0`

`clust.csh / clust-sil.csh`

使用目的 `clust` を基にして作った CSH プログラム。`clust-sil.csh` は 10 混合、1 状態の無音モデルを付加したものに对应。(距離計算のとき第 1 状態を無視)

使用法 `clust.csh` 希望するクラスタリング数 condition リスト
 condition リストを作成する必要がある。(クラスタリング対象のモデルを列挙する)

使用例 `clust-sil.csh 5 ../List/condition.285`

以下のような出力結果が得られる。

```
class = 0 center = 172 dist = 6240.775040 num. of members = 59
class = 1 center = 102 dist = 7240.285920 num. of members = 77
class = 2 center = 252 dist = 2094.782730 num. of members = 22
class = 3 center = 43 dist = 5536.106710 num. of members = 64
class = 4 center = 154 dist = 6133.889060 num. of members = 63
number of classes = 5
HMnet 0 belong 1
HMnet 1 belong 1
HMnet 2 belong 1
```

2.5.3 belong.csh

使用目的 `clust.csh / clust-sil.csh` の出力ファイルをもとに、中心モデルとクラスタメンバーのリストを作成するシェルプログラム。

使用法 まず `clust.csh / clust-sil.csh` により、出力リストを得る。これを `clust.txt` とする。また `clust.csh / clust-sil.csh` で用いた condition リストのファイル名を `condition` とする。このとき以下を実行。

```
belong.csh clust.txt condition
```

結果の例を示す。先頭が中心モデル名、続いてクラスタメンバーが出力される。

```
F104 F101 F102 F205 F103 F403 F104 F105 F106 F107 F108
M101 M010 M101 M103 M105 M106 M108 M208 M401 M404
M204 M204 M406
```


全ファイルのリスト

```
CCLex/
Clustexe/
List/
Model/

CCLexe:
Hirei_Average_HMnet-nosil.csh*
Wave2Para.csh*
gousei2.csh*
link.csh*
makesample0.awk*
memory.txt
mklist6.csh*
nopau2.csh*
test.csh*

Clustexe:
belong.csh*
clust*
clust-init*
clust-init-sil*
clust-sil*
clust-sil.csh*
clust-table*
clust-thr*
clust.csh*

List:
condition.285
member.clust.me015mx1
member.clust.me015mx10
member.clust.me015mx15
member.clust.me015mx5
member.clust.me285mx10
member.clust.me285mx15
member.clust.me285mx2
member.clust.me285mx20
member.clust.me285mx5

Model:
FTK/
HMM.-.10mix
HMnet_filled.200.female@
HMnet_filled.200.male@
MHT/
MULTI_without_FYOMA

Model/FTK:
HMnet_filled.200
HMnet_filled.200.log
HMnet_filled.300
HMnet_filled.300.log
HMnet_filled.400
HMnet_filled.400.log
HMnet_filled.600
HMnet_filled.600.log

Model/MHT:
HMnet_filled.200
HMnet_filled.200.log
HMnet_filled.300
HMnet_filled.300.log
HMnet_filled.400
HMnet_filled.400.log
HMnet_filled.600
HMnet_filled.600.log
```

参考文献

- [1] 鷹見 淳一, 嵯峨山 茂樹: “音素コンテキストと時間に関する逐次状態分割による隠れマルコフ網の自動生成”, 音声研資, SP91-88(1991.12).
- [2] 鷹見 淳一, 嵯峨山 茂樹: “隠れマルコフ網 (HM-Net) を用いた話者適応”, 音講論, 1-1-8(1992.3).
- [3] 小坂 哲夫, 鷹見 淳一, 嵯峨山 茂樹: “話者混合 SSS による不特定話者音声認識と話者適応,” 電子情報通信学会技術研究報告, SP92-52, pp. 17-24 (1992.09).
- [4] 小坂 哲夫, 松永 昭一, 倉岡 幹雄: “話者クラスタリング手法を用いた不特定話者音素モデルの作成,” 電子情報通信学会技術研究報告, SP94-80, pp. 9-16 (1995.01).