音声言語情報処理　4－13
（1994. 12. 16）

非 left - right HMMの性質に関する検討

松尾　広　　石亀　昌明
秋田大学鉱山学部情報工学科
〒010　秋田市手形学園町 1 - 1

あらまし

本報告では，非 left - right HMMにおける状態遷移が，ほとんど left to right になるという性質に注目し，従来の left - right HMMとの比較検討を行なった．音素認識実験を行ない，学習済みの非 left - right HMMを left to right 化することができるが，left - right HMMでは，適切な初期値を選んでも，left to right 化した非 left - right HMMに匹敵するパラメタを学習できないことがわかった．その結果，学習時に非 left - right の構造を持っていることが重要であることが示された．

和文キーワード　HMM　音素認識　非 left - right HMM　学習

A Study on the Characteristic of Non Left - Right HMM

Hiroshi Matsuo and Masaaki Ishigame
Department of Information Engineering, Mining College, Akita University
1 -1 Tegata Gakuen-machi, Akita-shi, 010 JAPAN

Abstract

This paper compares non left-right HMM with conventional left-right HMM taking notice of the characteristic that state transitions of non left-right HMM seems almost left-to-right. From phoneme recognition experiments, it is shown that trained non left-right HMMs can be translated into left-right HMMs but left-right HMMs can not get parameters showing comparable performance to the translated HMMs even if choosing initial parameters carefully. This result shows importance of non left-right HMM structure when training parameters.

英文 key words　HMM, phoneme recognition, non left-right HMM, training

1. はじめに

HMMを用いた音声認識において，HMMの精密化は，認識システムの性能向上に必要である．HMMのトポロジー（状態数と状態遷移の型）は，HMMの性能を決める要因のひとつであるが，パラメタの学習をする前に決めておかなければならないにもかかわらず，先見的に決定する方法がない．そのため，いくつかのトポロジーを与えておいて，実験的に決める方法[1]や，混合分布をコンテキスト方向／時間方向に分割することによって，逐次的に決定する方法[2]がとられている．筆者らは，状態遷移の型は状態遷移確率で表されうると仮定し，トポロジーを学習的に決定する方法として，音素内の状態遷移がエルゴード的，音素間の状態遷移がleft to rightである非left - right HMMを用いる方法を提案，音素認識実験から，その有効性を示した[3]．この方法は，状態遷移の制限が少なくなったほかは，従来のHMMと変わりがないので，同じ学習方法／手順がそのまま適用可能である．

本報告では，非left - right HMMにおける状態遷移が，ほとんどleft to rightになるという性質に注目し，非left - right HMMと従来のleft - right HMMとの比較検討を行なった．特に，非left - right HMMは，学習の結果，left to rightになるのに，left - right HMMでは非left - right HMMに匹敵する性能が得られない原因として，学習時に非left - rightの構造を持っていることが重要であることを，音素認識実験により，実験的に示す．

2. 非 left - right HMMによる音素認識

非left - right HMMの構造を図1に示す．非left - right HMMは，音素内の状態遷移はエルゴード的，音素間の状態遷移はleft to rightである．2つの音素モデルの接続に関して，前後の音素モデルの状態間それぞれに状態遷移のパスを持つものを音素間多重結合，音素間の状態遷移を1つだけ持つもの（あるいは，音素内に最終状態を持つもの）を音素間単結合と呼ぶ．音素間多重結合を持つ場合のトレリスを，図2に示す．

単語音声（東北大－松下単語音声データベース，212単語×男女各10名，25135音素）を用いた音素認識実験の結果，コンテキスト依存の場合，非left - right HMMは従来のleft - right HMMをうわまわる音素認識率が得られることが示された[3]．

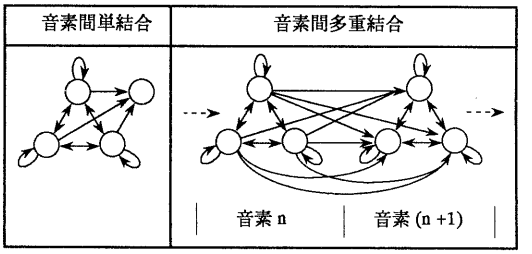

図1　非 left - right HMM

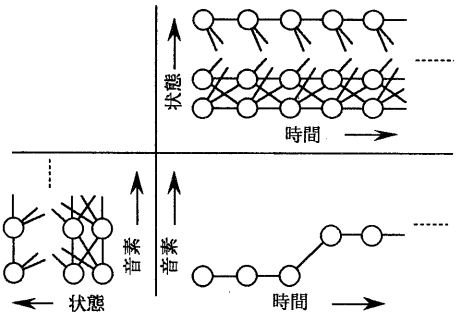

図2　音素間多重結合を持つ
非 left - right HMMのトレリス

3. 非 left - right HMMの性質について

いくつかの追加実験の結果から，次のような非left - right HMMの性質がわかった．

(1)状態遷移確率は有効であり，トポロジーを決めるのに役立っている．

音素認識実験では，出力確率分布として連続分布を用いているため，状態遷移確率は出力確率分布に対してダイナミックレンジが小さく，

状態遷移確率が有効に働いていないことが考えられないではない．しかしながら，12状態の単一分布型非 left - right HMMと１状態12混合数の混合分布型HMMを比較する（共に出力確率分布数は12）と，非 left - right HMMの方が認識率が高く，状態遷移には意味があることがわかる[4]．

(2)非 left - right HMMはパラメタ数が多いが，パラメタ数の多さが非 left - right HMMの本質ではない．

12状態単一分布型 left - right HMMと，４状態３混合数の混合分布型非 left - right HMMを比較すると，出力確率分布数はどちらも12で同じであるが，状態遷移確率のパラメタ数の比は222:32（音素間多重結合の場合）で前者の方が約７倍多い．しかしながら，音素認識率は後者の方が高い．また，12状態単一分布型非 left - right HMMの方が，さらに状態遷移確率のパラメタ数が多いのにもかかわらず，音素認識率が高いので，学習サンプル不足ではないと考えられる．パラメタ数の多さのためだけに音素認識率が高くなるとはいえないことを示している[3]．

(3)状態遷移の原因が与えられていれば，非 left - right HMMは正しくトポロジーを学習できる．

音素認識実験から，コンテキスト依存ならば非 left - right HMMの方が音素認識率が高く，非 left - right HMMにはコンテキスト情報が必要であることが示された[3]．

コンテキストによって性質が変化する信号源を使ってシミュレーションを行なってみると，コンテキスト情報を与えた場合（モデルをコンテキスト依存にした場合），非 left - right HMMは期待したとおりのパラメタを学習することができた[5]．そのとき，left - right HMMよりも正しいパラメタに到達できる初期値の範囲が広かった．これは，コンテキスト依存にする必要があるという音素認識実験の結果と一致し，left - right HMMは初期値の影響を受けている可能性も考えられる．

(4)非 left - right HMMは連結学習を用いてパラメタの学習を行なうが，left - right HMMに比べて学習区間の影響を受けにくい．

非 left - right HMMは連結学習を用いてパラメタの学習を行なうが，視察によるラベルを用いて音素の学習区間を限定していた[3]．ラベルによる音素境界に対して，前後にマージンを設けて学習区間を広げた場合，非 left - right HMMでは認識率の低下がほとんどないのに対して，left - right HMMでは認識率が低下する[6]．非 left - right HMMは，ある程度の学習区間の広がりに対しても安定である．

(5)非 left - right HMMの実際の状態遷移は，left to right に非常に近い．

Viterbi アルゴリズムを使って，音素内の状態遷移系列を見ると，音素内の２状態間の状態遷移が両方向に起きてもよいようになっているにもかかわらず，ほとんどが片方向に集中している（表１）．また，ある状態から別な状態を経由して，再び同じ状態に戻ってくる，状態の循環もほとんどない[3]．２つのことから，非 left - right HMMの音素内の状態遷移は，ほとんど left to right であることがわかる．

表１　非 left - right HMM における状態遷移の偏り

バランス	0.5	0.5～0.7	0.7～0.9	0.9～1.0	1.0
学習時	0.6	1.5	2.7	4.4	90.8
認識時	0.4	0.4	0.7	0.4	98.1

すべての状態の対に対する割合（%）

バランス（状態遷移の偏りの大きさ）

$$= \frac{\max(n[i,j], n[j,i])}{n[i,j]+n[j,i]}$$

n[i,j]：状態iからjへ遷移した回数

状態の循環	無	有
学習時	98.8	1.2
認識時	99.0	1.0

該当する音素の割合（%）

4. 非 left - right HMMと left - right HMMの違い

コンテキスト依存の非left - right HMMは，学習の結果，ほぼleft to rightになることが示された．それならば，left - right HMMで非left - right HMMに匹敵するパラメタが学習できないのは，何が原因であろうか．原因の一つとして，シミュレーションの結果で示されるような，初期値の影響が考えられる．これを確かめるために，以下の実験を行なった（実験条件を表2に示す）．

表2　実験条件

音声資料	東北大－松下単語音声データベース 212単語，男女各10名 25カテゴリ，25135音素
分析条件	29ch BPF，10ms/frame フレームごとにK-L展開で，10次元に圧縮後， 5フレーム50次元をK-L展開で10次元に圧縮 したものを特徴パラメタとする
HMM	12状態，出力確率分布は状態ごと持つ 連続分布（ガウス分布，単一分布） 音素間多重結合 left - right HMMは，左への遷移のみ許さない

4.1 非 left - right HMMの left - right 化

まず，非left - right HMMにおける状態遷移がleft to rightになっていることを確かめるために，学習済みの非left - right HMMをleft - right HMMに変換することを試みた．

非left - right HMMをleft - right化するために，音素内の状態遷移数の行列から，状態遷移ができるだけleft to rightに近づくように，状態を並べ換える．音素内の2状態間の状態遷移がほとんど片方向であって，もう一方向は無視できると仮定して，状態遷移確率を0においてleft - right化を行なう．以下に，状態並べ換えの手順とleft - right化の手順を示す．並べ換えを適用した例を表3に示す．

[並べ換え手順]

```
for i = 1 to N
  order'[i] = i
f' = ev( order' )
repeat
  f = f'
  order = order'
  f' = -∞
  for i = 1 to N
    for j = 1 to N
      if i = j continue
      order" = mv( order, i , j )
      if ev( order" ) > f' then
        f' = ev( order" )
        order' = order"
      endif
    next j
  next i
while f' > f
```

ここで，

N：状態数

order，order'，order"：並べ換えられた状態（orderに結果が入る）

order[i]：並べ換えられた状態のi番目

mv(order, i , j)：並べ換え関数

i番目の要素を取り出し，一旦詰めた後，j番目の要素の前に挿入する．

（例）**mv**((1, 2, 3, 4, 5), 2, 4)
→ (1, 3, 4, 2, 5)

ev(order)：評価関数

ev(order)

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \{ n[order[i],order[j]] * \mathbf{sign}(j-i) \}$$

n[i,j]：音素内の状態iからjへ遷移した回数

$$\mathbf{sign}(x) = \begin{cases} -1 & : x<0 \\ 0 & : x=0 \\ 1 & : x>0 \end{cases}$$

[left - right 化手順]

```
for i = 1 to N
  for j = 1 to N
    if j < i then
      p[ order[i], order[j] ] = 0
    endif
  next j
next i
```

ここで,

p[i, j]:音素内の状態 i から j へ遷移する確率

上記の手順によって状態を並べ換え，left - right 化したパラメタを用いて音素認識実験を行なった．結果を表4に示す．非 left - right HMMは，若干音素認識率が低下する程度で left - right HMM に変換可能であり，状態遷移がほとんど left to right であることが示された．

表4　left - right HMM 化したときの音素認識率

非 left - right HMM	91.2
left - right 化後	90.2
left - right HMM	86.9

4.2 初期値の影響

シミュレーションの結果は，left - right HMM は初期値の影響を受けている（局所解に陥っている）可能性があることを示唆している．以下の実験では，パラメタの初期値を変えて学習したときの，音素認識率の変化を調べることにより，初期値の影響について検討を行なった．

表3　状態並べ替えの例 /a/

(a) 音素内の状態遷移回数

from \ to	0	1	2	3	4	5	6	7	8	9	10	11
0	789	1	3	449	0	52	0	0	11	0	17	30
1	0	749	0	1	1014	0	359	0	8	12	1	0
2	0	0	1022	0	0	0	0	0	562	0	0	5
3	0	7	6	255	305	0	258	31	0	217	0	0
4	0	0	0	0	7248	0	0	2	3	1331	0	0
5	0	5	23	45	0	876	0	43	0	26	0	2
6	3	0	5	1	18	2	4338	0	0	371	0	3
7	0	0	713	2	2	0	0	424	0	0	1	1
8	131	2	0	5	0	2	0	3	1972	4	0	1
9	3	0	303	0	0	0	0	1301	3	769	0	2
10	0	490	1	96	0	6	0	0	9	0	459	12
11	2	194	19	5	0	4	0	8	20	0	0	164

(b) 状態並べ替え後

from \ to	0	10	11	5	3	1	6	4	9	7	2	8
0	789	17	30	52	449	1	0	0	0	0	3	11
10	0	459	12	6	96	490	0	0	0	0	1	9
11	2	0	164	4	5	194	0	0	0	8	19	20
5	0	0	2	876	45	5	0	0	26	43	23	0
3	0	0	0	0	255	7	258	305	217	31	6	0
1	0	1	0	0	1	749	359	1014	12	0	0	8
6	3	0	3	2	1	0	4338	18	371	0	5	0
4	0	0	0	0	0	0	0	7248	1331	2	0	3
9	3	0	2	0	0	0	0	0	769	1301	303	3
7	0	1	1	0	2	0	0	2	0	424	713	0
2	0	0	5	0	0	0	0	0	0	0	1022	562
8	131	0	1	2	5	2	0	0	4	3	0	1972

4.2.1 初期値の設定

非 left - right HMMの初期パラメタの設定に関しては，有声破裂音の認識実験の際に検討を行なっている[7]．その結果，乱数を用いて初期化するのがよいという結果が得られている．本報告では，全音素の認識において，さらにいくつかの方法で非 left - right HMMと left - right HMMの初期値を設定して，パラメタの学習を行なった．

［初期値設定法］

特に断りがないとき，共分散行列の初期値は単位行列，状態遷移確率は定数（必要なところには0をおく）とする．

A. 乱数によって平均ベクトルを設定

今までの認識実験で用いている設定法．乱数によって設定される平均ベクトルは，音声スペクトル集合からは遠い．この方法で学習されたパラメタを設定法C～Fで用いる．

B. 音声スペクトル集合をK-means クラスタリングし，平均ベクトルを設定

left - right HMMで用いることを考慮し，クラスタの系列からクラスタ間の遷移行列を求めて，**4.1** で述べた方法を用いてクラスタを並べ換えておく．

C. 学習済みの left - right HMMの出力確率分布から平均ベクトルを移植

D. left - right 化した学習済みの非 left - right HMMの出力確率分布から平均ベクトルを移植

E. 学習済みの left - right HMMの出力確率分布を移植

状態遷移確率のみを学習する．

F. left - right 化した学習済みの非 left - right HMMの出力確率分布を移植

状態遷移確率のみを学習する．

4.2.2 音素認識実験

上記の初期値設定法を用いたときの音素認識率を，図3に示す．設定法Fでは left - right HMMは，パラメタの学習ができなかった．

非 left - right HMM，left - right HMMのいづれも，BよりもAの方が高い認識率が得られた．非 left - right HMMは非 left - right HMM，left - right HMMは left - right HMMに由来する初期値を設定したとき，元のモデルに近い認識率を示す．また，left - right HMMに由来する初期値を用いた非 left - right HMM（C，E）では，認識率は元の left - right HMMより若干良い程度であり，Aの非 left - right HMMに及ばない．left - right 化した学習済みの非 left - right HMMに由来する初期値を用いた left - right HMM（D，F）では，Aの非 left - right HMM，および**4.1** の left - right 化した非 left - right HMMに及ばず，Aの left - right HMMと同程度の認識率しか得られなかった．

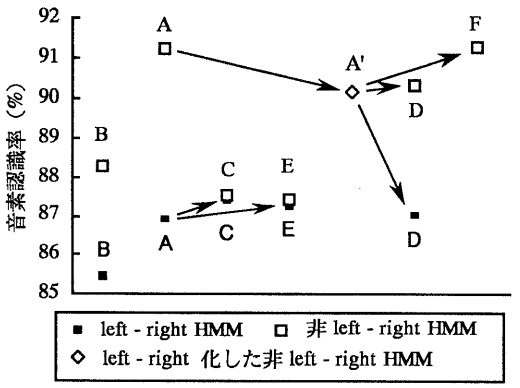

図3　初期値設定法による音素認識率の差

5. 考察

初期値の影響を調べた実験から，非 left - right HMM，left - right HMMのいづれも，乱数で初期化したときの認識率が，それぞれのほぼ上限になっている．設定法Bの，音声スペクトル集合に属する，妥当そうな初期値よりも，乱数を用いたほうが認識率が高い．また，設定法C～Fを用いた場合は，元のモデルの認識率を大きく越えることはなかった．これらは，音声スペクトル集合に属するような平均ベクトルの初期値を設定した場合，そこに強く引き込まれてしまうためと考えられる．初期値の設定法を工夫しても，認識率は劇的には向上しないことから，乱数による初期値で得られたパラメタによる認識率は，学習で得られる認識率の上限であり，シミュレーションの結果から懸念されたような，初期値の影響によって局所解に陥っていた可能性はほとんどないものと思われる．

注目すべきなのは，left - right HMMでは，パラメタとしてはとりうるのに，学習では獲得できないことがある，という点である．非 left - right HMMの left - right 化で示したように，非 left - right HMMに近い性能を示す left - right HMMは存在する．しかし，left - right 化した非 left - right HMMと同等の性能を持つパラメタを最初から left - right HMMに学習させるために，もっともらしい初期値を設定しても（**4.2.1** 設定法D，F），期待どおりのパラメタを学習することができなかった．これは，学習の結果として left to right になるとしても，学習の際には非 left - right HMMの構造を持っていることが重要であることを示していると考えられる．

今までの実験結果と，本報告における実験結果は，非 left - right HMMのパラメタの学習能力が高さの現われであると思われる．音素内のトポロジーの学習に限らず，連結学習において，学習区間の広がりに対して追従できる[6]柔軟性も，学習能力の高さの現われといえよう．

6. まとめ

非 left - right HMMにおける状態遷移が，ほとんど left to right であるという性質に注目し，非 left - right HMMと left - right HMMの比較検討を行なった．

まず，学習された非 left - right HMMがほとんど left to right であることを確認するため，非 left - right HMMの left - right 化を行ない，若干認識率が低下する程度で，変換可能であることがわかった．

次に，非 left - right HMMが left - right 化できるのに，left - right HMMではそれに匹敵するパラメタが学習できない原因について検討した．シミュレーションでは初期値によって局所解に陥ってしまう可能性が見られたので，いくつかの方法で初期値の設定を行ない，音素認識率の変化から初期値の影響を調べた．その結果，初期値の設定法を変えても，乱数で初期化した場合を大きく越えないことから，局所解に陥っていた可能性はほとんどないと考えられる．

一方で，left - right 化した非 left - right HMMと同等の性能を持つパラメタを，最初から left - right HMMに学習させることに失敗したことから，パラメタとしてはとりうるのに，学習できない場合があることがわかった．これは，学習の際に，非 left - right HMMの構造を持っていることが，重要であることを示している．

謝辞

この研究の1部は，文部省科学研究費の補助を受けた．

参考文献

(1)K-F. Lee : "AUTOMATIC SPEECH RECOGNITION The Development of the SPHINX system" Kluwer Academic Publishers pp. 82-83 (1989)
(2)鷹見，嵯峨山："逐次状態分割法による隠れマルコフ網の自動生成"信学論 vol.J76-D-Ⅱ pp.2155-2164（1993-10）
(3)松尾，石亀："トポロジーの学習と音素モデル間の相互依存を考慮したHMMによる音素認識"信学論 vol.J76-D-Ⅱ pp.1835-1842（1993-9）
(4)松尾，石亀："非 left - right HMMの構造に関する検討"音響講論 2-Q-7 (1993-10)
(5)松尾，石亀："シミュレーションによる非 left - right HMMの特性の検討"音響講論 2-P-16 (1994-3)
(6)松尾，石亀："非 left - right HMMにおける学習／認識区間の影響"音響講論 1-R-5 (1994-10)
(7)松尾，石亀："非 left - right HMMによる有声破裂音の認識"音響講論 2-1-2 (1992-3)