HMM の構造探索による音素モデルの生成

池田 思朗

東京大学 計数工学科

〒 113　東京都 文京区 本郷 7-3-1　email: shiro@bcl.t.u-tokyo.ac.jp

HMM により音素モデルを構成する場合，その構造は試行錯誤的に知識や経験によって決定されることが多い．これに対し，パラメタ数とデータ数を考慮した情報量基準を用い，これを最小化するようにモデルの構造を変化させ，最適化するアルゴリズムを提案する．構造を変化させる毎にパラメタを推定し，尤度を最大化するように，状態を分割あるいは新しく状態遷移を定義することにする．求まったモデルのパラメタを再度推定し，情報量基準 AIC でそのモデルを評価し，さらに構造を変化させるか否かを決定する．このアルゴリズムを用いて，モデルを構成し，計算機上の実験を行ない，さらに音素認識実験を行なった結果を示す．

隠れマルコフモデル，音素モデル，構造探索，赤池情報量基準

Construction of Phone HMM Using Model Search Method

Shiro Ikeda

Department of Mathematical Engineering and Information Physics, University of Tokyo

7-3-1, Hongo, Bunkyo-ku, Tokyo, 113, JAPAN　email: shiro@bcl.t.u-tokyo.ac.jp

In most of the conventional approaches of speech recognition using HMM, each model is chosen according to the heuristic knowledge. In this technical report, a new algorithm to construct a HMM automatically is given. The algorithm searches the model which would minimize the AIC (Akaike's information criterion) by increasing the states or the connections between the states. AIC considers not only the likelihood function but also the number of the parameters and the training data. Two recognition experiments using the models constructed with this algorithm is shown. First, artificial data are used and second, ATR speech database are used.

HMM, phone models, model search, AIC

1 はじめに

HMM(隠れマルコフモデル) を用いて音素や単語を表し，これによって音声認識システムを構築するアプローチが最近盛んに行なわれている [1]．このときの一つの大きな問題は，状態数や状態遷移によって定義される HMM の構造をいかに決めるかである．実際に存在するシステムを見ても，決め方はそれぞれによって異なっているが，多くの場合は試行錯誤的に経験をもとに決めていることが多い．

HMM の構造決定の問題は，確率モデルの構造をいかに決めるかということである．この問題に対しては，統計の分野において AIC (赤池情報量基準) を始めとするいくつかの情報量基準が提案されている (**2** 節)．これらを用いてモデルの良さを測ることによって，この問題は解決できるはずである．AIC はそのパラメタを最尤推定によって求めたモデルに対して用いることができ，Baum-Welch (B-W) アルゴリズムを用いてパラメタを推定した HMM に対して用いることができる．これが本報告の出発点である．

ここで問題となるのは，AIC 等の情報量基準を用いる場合，一つのカテゴリに対して構造の異なった複数のモデルを用意しなければならない点である．各モデルのパラメタを推定するのには B-W アルゴリズムを行なう必要があり，時間的な制約から，用意できるモデルには制限がある．

特に HMM のような複雑なモデルでは，多くの構造が考えられる．では，複数のモデルを用意しておくのではなく，モデルの構造を変化させていくによって，より良いモデルの構造を探索していくことは考えられないだろうか．実は AIC を用いる場合，比べられるモデルはその構造が層構造になっていることが必要である．すなわち，複雑なモデルはより単純な構造のモデルを含んだ形になっている必要がある．これを満たしながらモデルを順に複雑にしていくことで，最適なモデルの構造を探索することができるのではないだろうか．

以上の考え方に基づき，HMM の構造探索アルゴリズムを提案する [2]．このアルゴリズムでは，1) 状態数あるいは状態遷移を増やしていくことでモデルを順に複雑にしていき，2) 複雑にしていく段階毎に，パラメタの推定を行ない，AIC でモデルを評価する．これを繰り返し行ない，AIC が最も小さくなるモデルを選ぶことにする．詳細については **3** 節で述べる．

提案するアルゴリズムを用いて行なった 2 つの実験について，**4** 節 で述べる．1 つ目の実験は，計算機内に定義した複数の隠れマルコフモデルを信号源としてデータを生成し，そのデータを認識するという実験である．もう 1 つは ATR 音声研究用データベースを用いての音素認識実験である．

また，モデルの構造を変化させていき，HMM を定義しようという試みとしては，ATR の鷹見らの行なった SSS(successive state splitting) と呼ばれる手法がある [3]．SSS では，時間方向と同時にコンテキスト方向に HMM を分解していくが，本研究で提案するアルゴリズムでは，ある一つのカテゴリに属するデータに対して，HMM を構成することが目的であり，カテゴリ方向への分解は行なわない．この点で SSS とは扱う問題が異なっている．実験の考察とともにこれ以上の説明は **5** 節で述べることにする．

2 AIC

以後，HMM としては出力確率分布として離散分布のものを考える．

モデル選択の基準として用いられる AIC は [4][5]，サンプルデータに対するモデルの当てはまりの良さを表す尤度の項とモデルの複雑さを表すパラメタ数の項によって定義されている．一般にモデルのパラメタが多くなれば，モデルは複雑になり，表現力が増す．これによってサンプルデータをより詳細に表すことができるのであるが，他のデータに対しては誤差が大きくなる可能性が増す．この 2 つのバランスをとることでモデルの良さを表そうというのが AIC である．

実際の分布 $p(y)$ とモデルの確率分布 $f(y|\theta)$(ただし θ はパラメタ) との間の “近さ” を Kullback 情報量，

$$\begin{aligned} D(p,f) &= \int p(y)\log\frac{p(y)}{f(y|\theta)}dy \\ &= \int p(y)\log p(y)dy - \int p(y)\log f(y|\theta)dy \end{aligned} \tag{1}$$

を用いて定義する．これを最小にするパラメタは明らかに最尤推定によって求められた θ^* であることから，以後は最尤推定点での振舞いを考える．式 1 の第 1 項はモデルと関係のない項なので無視し，第 2 項のみを考える．最尤推定されたパラメタ θ^* は訓練用データから推定された点であることから，その選び方によって真の値 θ_o を中心に分布する．データの確率分布を考慮に入れ，$D(p,f)$ の期待値を考える．訓練用データのセットを Y とし，

$$E_p^Y[D(p,f)] \tag{2}$$

を求めたいのだが，真の分布 p は分からない．これを訓練用データから推定し，その推定値の $2\times n$(訓練用データ数) 倍を AIC として定義する．これはパラメタ数を m とすると，

$$\mathrm{AIC} = (-2)\sum_{i=1}^{n}\ln f(y_i|\theta^*) + 2m \tag{3}$$

となる．HMM に対して AIC を用いる場合，HMM の全てのパラメタが独立でないことを考慮に入れなければならない．AIC は最尤推定点のまわりで，

$$\left.\frac{\partial f(y|\theta)}{\partial \theta_i}\right|_{\theta^*} = 0 \qquad for\ \forall\ i \qquad (4)$$

が成り立つとして定義されている．B-W アルゴリズムの収束点として求められたパラメタでは全てが独立でないので，式 4 は成り立たない．ただし，独立でないパラメタを取り除くことによって，この問題は解決できる．状態遷移確 $\{a_{ij}\}$ を例にとると，ある 0 でない a_{ij} に対し，

$$a_{ij} = 1 - \sum_{j':j'\neq j} a_{ij'}$$

として消去すれば良い．このとき，AIC は同様に定義できる．独立でないパラメタは，離散分布を出力確率分布として持つ HMM の場合，丁度状態数の 2 倍個ある．AIC は，状態数を s として次のようになる．

$$\mathrm{AIC} = (-2)\sum_{i=1}^{n} \log f(y_i|\theta^*) + 2(m-2s) \qquad (5)$$

3 提案するアルゴリズム

3.1 状態数を増やす

HMM 構造を決めるということは，その状態数と状態遷移をどう決めるかということである．まず，状態を一つづつ増やしていくことによって状態数を決めることができないかと考えてみる．そのとき，ランダムに状態を増やすのではなく，効率良く増やしていけることが望ましい．新しく増やす状態をどの場所に定義するか，そのときの初期値はどうするか，どこまで増やすか，の 3 つの問題について順に見ていく．

被分割状態の決定

どの状態を分割するかの決定には，その状態の出力確率分布と，その状態にどのくらいの時間留まるかの期待値を見て決めれば良いと考えられる．出力確率分布は，そこで受けとることのできるシンボルを反映する．これがばらついている場合，そのエントロピーも増えるはずである．したがってその状態の悪さとしてエントロピーを用いることができる．また，ある状態に長くいることを避けるために，そのエントロピーにある状態に留まる時間の期待値を乗じることにした．すなわち，

状態 i に留まる時間の期待値

$$\propto \sum_{t=1}^{T-1}\sum_{j=1}^{N} \alpha_i(t)a_{ij}b_j(y_{t+1})\beta_j(t+1) \qquad (6)$$

状態 i のエントロピー

$$= \sum_{k}^{K} b_i(k)\log b_i(k) \qquad (7)$$

被分割状態の決定の際の比較量

$$= \sum_{t=1}^{T-1}\sum_{j=1}^{N} \alpha_i(t)a_{ij}b_j(y_{t+1})\beta_j(t+1) \times \sum_{k} b_i(k)\log b_i(k). \qquad (8)$$

として，この値が最大となる状態を分割することにする．

初期値の設定

被分割状態を決めた後，どのように分割し，それぞれのパラメータの初期値を決定するかが問題となる．

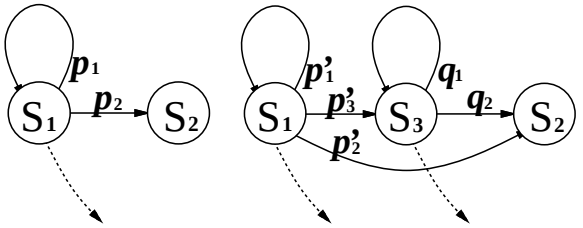

図 1: 状態分割

図 1 の 2 つのモデルを考える．左側の図に於いて，状態 S_1 に n 回留まり，次の時間に S_2 にいる確率は確率 P_n は $p+q \leq 1$ に注意して，

$$P_n = p_1^n p_2 \qquad (9)$$

となる．一方これに状態 S_3 を挿入した右側の図ではこれは式 10 のようになる．

$$\begin{aligned} P'_n &= p'^n_1 p'^n_2 + p'^n_3 q_2^n \sum_{i=0}^{n-1} p'^i_1 q_1^{n-i-1} \\ &= \begin{cases} p'^n_1 p'_2 + q_2 p'_3 \dfrac{p'^n_1 - q_1^n}{p'_1 - q_1} & p'_1 \neq q_1 \\ p'^n_1 p'_2 + n q_2 p'_3 p'^{n-1}_1 & p'_1 = q_1 \end{cases} \end{aligned} \qquad (10)$$

ここでもし，$p'_2 = p_2$, $q_1 = p_1$, $q_2 = p_2$, $p'_1 + p'_3 = p_1$ ならば，式 10 の P'_n は式 9 の P_n と等しくなる．したがって，これらの条件を満たした上で状態 S_1 と S_3 の出力確率分布が等しければ，2 つのモデルは外から見る限り全く等しくなる．つまり，最尤推定されたモデルと全く等価なモデルを初期値とすることができる．もし，分割する前のモデルが完全にサンプルデータを表現しているのならば，パラメタを推定しても尤度は上がらないが，そうでないならば，構造を複雑にすることで，より良いモデルへになる．また，このように状態を定義していくと，AIC を用いる場合に問題になる階層構造の条件を満たしていることがわかる．

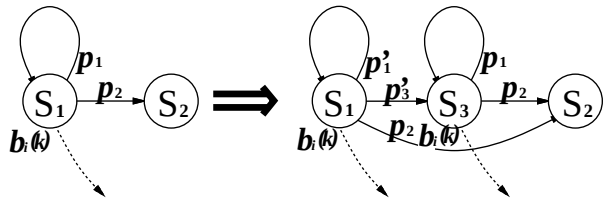

図 2: 状態分割の初期値

どこまで状態を増やすか

これには **2** 節で述べた AIC を用いることにする．状態を増やしていっても AIC が減らない場合，それ以上状態を増やすのを止めることにする．

以上をまとめて，状態を増やしていく手順は次のようになる．

1. 被分割状態を決める．
2. 初期値を与える．
3. B-W アルゴリズムを行ない，最尤推定点を求める．
4. AIC を計算し，前の時よりも良くなれば 1 へ戻る．そうでない場合はそこで終了する．

3.2 状態遷移を増やす

状態の数が決定された後，どの状態からどの状態への状態遷移を定義するかという問題がある．状態遷移も段々に増やしていくことで，より良いモデルを推定できないだろうか．

この場合，状態そのものを増やした場合と異なり，新たな状態遷移を定義しながらも定義する前と等価なモデルを作るということができない．つまり，状態遷移を増やすということは全く違うモデルへ移るということになる．また，それによって得られるモデルが実際にうまくいっているのかは，B-W アルゴリズムを行なわなければ分らない．

一つ一つの状態遷移を新しく加え，B-W アルゴリズムを行なうことは時間的に不可能である．そこで，加えることによって最も尤度が上がっていきそうな状態遷移を予測し，その状態遷移を新しく加えることにする．予測には，B-W アルゴリズムの 1 ステップでどのくらい尤度が増すかを用いて決定したいのだが，ここでもまた，候補の数が多く，それぞれに 1 度づつ B-W アルゴリズムを行なうのは時間がかかる．そこで，B-W アルゴリズムの 1 ステップでの各状態遷移確率の増加分 $(\hat{\theta}-\theta)$ と $\partial \log P(y|\theta)/\partial\theta$ との積でこれを置き換え，この値によって新しく定義する状態遷移を決定する．この量は一回の B-W アルゴリズムで，全ての候補に対して計算できる．B-W アルゴリズムの 1 ステップでの状態遷移確率の増加分と尤度の変化分について順に見ていく．

状態遷移確率の増加分

まず，前提となる $\{a_{ij}\}$ に関する B-W アルゴリズム，式 11 を見てみる．

$$\hat{a}_{ij} = \frac{a_{ij}\dfrac{\partial \log P(y|\theta)}{\partial a_{ij}}}{\displaystyle\sum_j a_{ij}\frac{\partial \log P(y|\theta)}{\partial a_{ij}}} \tag{11}$$

B-W アルゴリズムの 1 ステップでどのくらい状態遷移確率が変わるかであるが，この式では，分母と分子それぞれに a_{ij} がかかっていることから，もし $a_{ij}=0$ ならばその状態遷移確率はこの更新ルールでは変化せず，したがって HMM の構造は変化しない．では，いままで 0 であった a_{ik} に小さな値 δ を入れ，新しく状態遷移確率を定義したとして新しく推定される $\hat{a}_{ij}$ がどのようになるかをみてみる．

$$\hat{a}_{ij} = \frac{a_{ij}\dfrac{\partial \log P(y|\theta)}{\partial a_{ij}}}{\displaystyle\sum_{j:j\neq k} a_{ij}\frac{\partial \log P(y|\theta)}{\partial a_{ij}} + \delta\frac{\partial \log P(y|\theta)}{\partial a_{ik}}},\ j\neq k \tag{12}$$

$$\hat{a}_{ik} = \frac{\delta\dfrac{\partial \log P(y|\theta)}{\partial a_{ik}}}{\displaystyle\sum_{j:j\neq k} a_{ij}\frac{\partial \log P(y|\theta)}{\partial a_{ij}} + \delta\frac{\partial \log P(y|\theta)}{\partial a_{ik}}} \tag{13}$$

$\partial \log P(y|\theta)/\partial a_{ik}$ は有限値をとり，発散することはない．式 13,12 より，もし δ が十分小さければその項は各式の分母の第 1 項に比べて十分小さいことになり，無視できる．したがって，a_{ij} $(j\neq k)$ に対しては，この条件の下でほとんど変化しないことになる．一方 a_{ik} の増加分であるが，これは以上の結果と，

$$\left.\frac{\partial \log P(y|\theta)}{\partial a_{ij}}\right|_{\theta^*} = \left.\frac{\partial \log P(y|\theta)}{\partial a_{ij'}}\right|_{\theta^*},\quad \text{for}\ \forall\ j'\ s.t.\ a_{ij'}\neq 0 \tag{14}$$

及び $\sum_j a_{ij}=1$ を考えて，

$$\begin{aligned}
\hat{a}_{ik}-a_{ik} &= \frac{\delta\dfrac{\partial \log P(y|\theta)}{\partial a_{ik}}}{\displaystyle\sum_{j:j\neq k} a_{ij}\frac{\partial \log P(y|\theta)}{\partial a_{ij}} + \delta\frac{\partial \log P(y|\theta)}{\partial a_{ik}}} - \delta \\
&\simeq \delta\left[\frac{\dfrac{\partial \log P(y|\theta)}{\partial a_{ik}}}{\displaystyle\sum_{j:j\neq k} a_{ij}\frac{\partial \log P(y|\theta)}{\partial a_{ij}}} - 1\right] \\
&= \delta\left[\frac{\dfrac{\partial \log P(y|\theta)}{\partial a_{ik}}}{\dfrac{\partial \log P(y|\theta)}{\partial a_{ij}}} - 1\right],\quad j\neq k\wedge a_{ij}\neq 0
\end{aligned} \tag{15}$$

とかける．$\{\pi_i\}$,$\{b_i(k)\}$ といった他のパラメータは δ が微小な場合にはほとんど変化しない．したがって，新しく状態遷移 a_{ik} を定義したときのパラメータ変化分は a_{ik} のみについて見れば良く，その値は式 15 から推定できる．これが正であれば，その状態遷移確率は増加する可能性があり，したがって新しく採用しても良いと考えられる．

尤度の増加分

一方，$\partial \log P(y|\theta)/\partial\theta$ はどうなるだろうか．a_{ik} を新しく定義するとき，これが十分小さければ，a_{ij} $(j \neq k)$，$\{b_i(k)\}$，$\{\pi_i\}$，がほとんど変化しないことから，$(\hat{\theta} - \theta)$ と $\partial \log P(y|\theta)/\partial\theta$ の積をとることを考えて，$\partial \log P(y|\theta)/\partial a_{ij}$ のみを考えれば良いことがわかる．これは B-W アルゴリズム，

$$\frac{\partial \log P(y|\theta)}{\partial a_{ij}} = \frac{\sum_{t=1}^{T-1} \alpha_i(t) b_j(y_{t+1}) \beta_j(t+1)}{\sum_i \alpha_i(T)} \tag{16}$$

から求めることができる．以上から，比較量として次式を得る．

$$\begin{aligned} \Delta_{ij} \log P(y|\theta) &\equiv (\hat{\theta} - \theta) \frac{\partial \log P(y|\theta)}{\partial \theta} \\ &\simeq \delta \left[\frac{\partial \log P(y|\theta)}{\partial a_{ik}} \Big/ \frac{\partial \log P(y|\theta)}{\partial a_{ij}} - 1 \right] \\ &\quad \times \frac{\sum_{t=1}^{T-1} \alpha_i(t) b_k(y_{t+1}) \beta_k(t+1)}{\sum_i \alpha_i(T)} \end{aligned} \tag{17}$$

このように定義された $\Delta_{ij} \log P(y|\theta)$ を候補に上がっている状態遷移それぞれについて求め，もっとも大きい値を示した状態遷移を採用することにする．

以上の方法では，全ての候補に対して $\Delta_{ij} \log P(y|\theta)$ を求めるのに B-W アルゴリズムを一回行なえば良い．

採用した後はそこに小さな値を代入し，B-W アルゴリズムを行なって新しいモデルを作ることにする．また，状態数を増加させた場合と同様に，AIC を用いて，どこまで状態遷移を増やすかを決めることにする．つまり，B-W アルゴリズムが収束する毎に AIC を計算し，AIC を最小にするモデルを最終的に採用することにする．

3.3 構造を変化させる

状態を増やす，あるいは，状態遷移を増やすアルゴリズムを組合せ，提案するアルゴリズムは次の通りである．

1. 状態数が 1 で，出力確率分布は一様分布の HMM を用意し，これを初期モデルとする．
2. 状態を増やし，B-W アルゴリズムでパラメータを推定する．
3. AIC の基準で AIC が減らなくなったらば 4 へ，そうでなければ 2 へ．
4. 状態推移を増やし，B-W アルゴリズムでパラメータを推定する．
5. AIC の基準で AIC が減らなくなったらば 2 へ，そうでなければ 4 へ．また，1 つも新しい状態遷移が定義されなかった場合は終了する．

4 実験

4.1 人工データを用いてのシミュレーション

ここでは，人工的に作った 5 つの隠れマルコフモデルを信号の発生源として用い，シミュレーションを行なった．5 つのマルコフモデルは，それぞれ状態数，状態遷移，出力確率分布のいずれかが異なっており．それぞれが出力する信号もこれに応じて異なったものとなっている．それぞれのマルコフモデルの出力した信号を 5 つのカテゴリとみなす．このとき，それらの信号がどの隠れマルコフモデルから発生されたかを認識するという問題を扱うことにする．発生源として用いたマルコフモデルを図 3 に示す．各モデルの各状態に定義された出力確率分布は離散分布で，6 つのシンボルを出力するものとする．

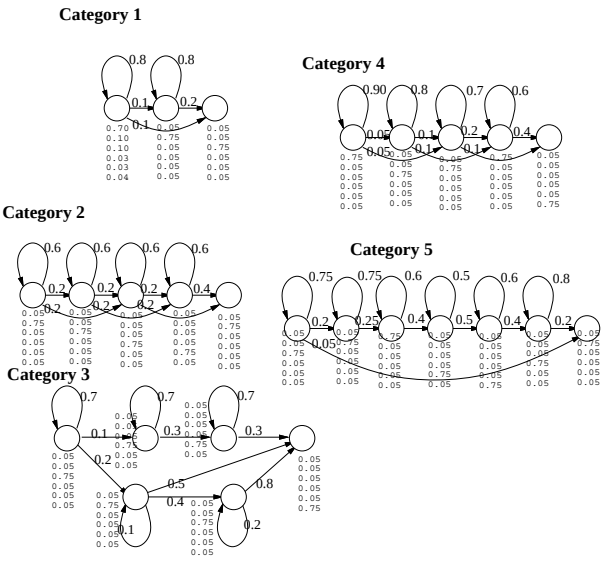

図 3: 信号発生用の HMM

実験の手順としては，次の通りである．

- 各モデルから，訓練用と認識実験用に 1000 個づつ計 2000 個の系列を発生させる．
- カテゴリの数だけモデルを作り，各データに対して構造とパラメータを推定する．
- 認識実験を行なう．コンテキストや文法の情報がないので，確率 $P(y|M_i)$ を最大にする M_i をその信号の属するカテゴリとする．

結果及び考察

結果として得られた HMM を，他のカテゴリに対してのモデルも含めて表 4 に示す．この図を見ると分かるが，

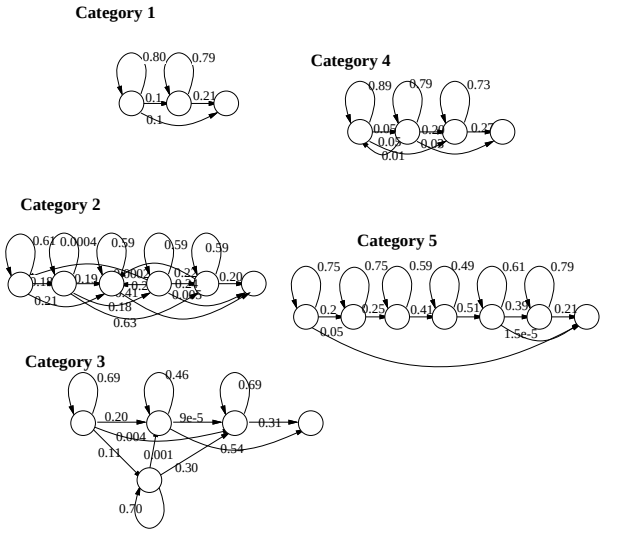

図 4: 構造を変化させた結果

信号発生用の隠れマルコフモデルの構造を良く反映しているものもあるが，そうではないものもある．

比較のために 3 状態，5 状態，7 状態の隠れマルコフモデルを用いて，同様の実験を行なった結果を上のアルゴリズムを用いた結果及び信号発生用のモデルを用いた結果とまとめて表 1 に示す．各カテゴリに対する認識率と，全体に対する認識率を示した．

表 1: 認識率

信号発生用	3 状態	5 状態	7 状態	構造探索
92.56%	84.92%	90.22%	90.68%	91.94%

このときの信号発生用モデル 3 が発生したデータに対する対数尤度を図 5 に示す．これを見ると，AIC を用いて構造探索したモデルが，認識用，訓練用データの双方に対して良いモデルであることがわかる．すなわち，データをよく表現していることがわかる．

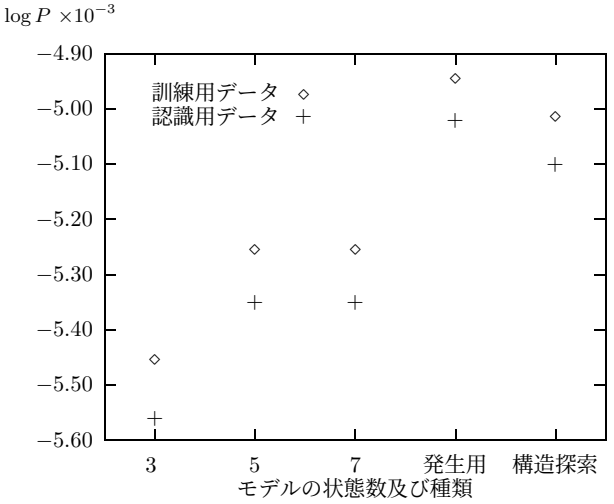

図 5: 用いたモデル間での尤度の比較

4.2 音声データを用いての実験

4.1 節の実験と同じアルゴリズムを用いて，音声データから切り出した音素に対し HMM を生成し，認識を行なった．実験は後続母音を/e/に限定し，(/m/, /s/, /t/, /k/, /d/) の 5 子音で行なった．

使用したデータ ATR 研究用日本語音声データベース [6] の男性話者 1 名 (MAU) によって発音されたデータであり，各音素データは音声データからラベルに従って切り出した [7].

ベクトル量子化用 タスクコード B,NA,NB,SY,F の 351 単語.

訓練用データ 単語発声データ (5240 単語) の偶数番目.

認識実験用データ 単語発声データ (5240 単語) の奇数番目.

認識タスク (/m/, /s/, /t/, /k/, /d/) の 5 子音．ただし，後続音素は e に限定した．

HMM の各状態の分布 離散分布，256 のシンボル.

シンボルの定義 logpow, mel-cep(15), Δlogpow, Δmel-cep(15) からなる 32 次元ベクトルを 256 にベクトル量子化 [8].

時間方向への状態分割制限 1 モデル当たりの状態数を最大 20 に制限.

分析条件 サンプリング周波数 20kHz,16bit 量子化，20msec ハミング窓，フレーム周期 5msec.

表 2: 構造探索によって求まったモデルの構造

子音 (訓練用データ数)	m (72)	k (107)	t (66)	s (75)	d (28)
状態数	11	20	20	13	9
状態遷移の数	32	71	64	40	27

認識タスクに選んだ 5 つの子音は，統計的処理をすることから，サンプルとして得られる数の多い 5 つの子音を用いた．

表 3 は，このアルゴリズムを適応させた結果のモデルと，比較のため，3,5,10,15,20 の状態数の隠れマルコフモデルを用いて行なった実験について，それぞれの誤認識率を示したものである．訓練用データと認識用データに対するそれぞれの値を示した．また，図 6 に子音/d/に対し，構造探索の結果得られたモデルを示す．

表 3: 誤認識率
(上段が認識用データに対するもの,
下段は訓練用データに対するもの)

3 状態	5 状態	10 状態	15 状態	20 状態	構造探索
10.3%	9.2%	8.0%	11.2%	14.4%	8.3%
(3.7%)	(2.0%)	(2.3%)	(1.4%)	(1.7%)	(2.0%)

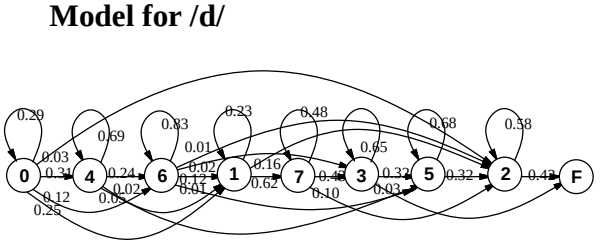

図 6: /d/に対しモデルを変化させた結果

表 3 を見る限り，提案するアルゴリズムを用いた場合と，全てを 10 状態で定義した場合は認識システムとして，ほぼ同じ振舞いを示している．しかし，訓練用，認識用データに対する尤度を見ると，これらの間には差がある．図 7 に/m/に対する各モデルの尤度を示す．

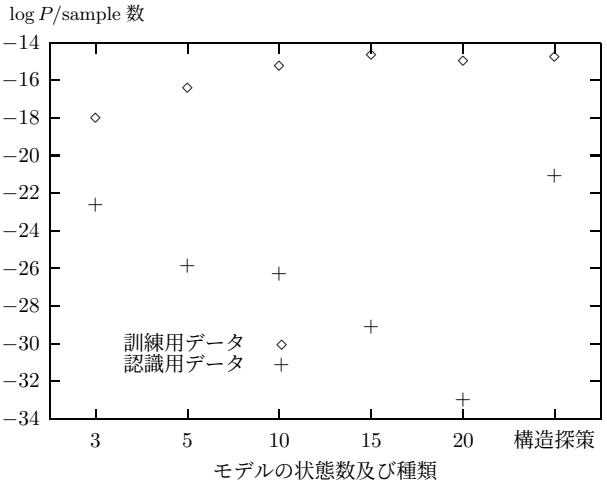

図 7: 用いたモデル間での尤度の比較

は構造を探索したモデルがどのモデルよりも良い値を示していることがわかる．これは，求めたモデルが他のモデルよりもデータを良く表現していることを示していると考えられる．

一般に，モデルの構造が複雑になれば，モデルの自由度が増え，訓練用データに対してはより良い尤度を示すようにパラメタが設定できるが，認識用データに対しての尤度は必ずしも良くならない．これを過学習と呼ぶが，この傾向が AIC を用いてモデル探索をしたことによってある程度押えられていることがわかる．ただし，このように AIC がうまく働くためには訓練用データの数が多くなければならず，ここには示されていないが，データの数が少なくなると，適切なモデルが推定できない場合がある．

これとは別に，HMM において B-W アルゴリズムを用いる場合，パラメタが極小解になってしまうという危険性がある．提案するアルゴリズムでの初期モデルは極小解が存在せず，B-W アルゴリズムによって推定されるパラメタは必ず最小解である．そこから階層的にモデルを複雑にしていくことから，極小解になる確率を減らせるのではないかという期待もできる．

また，AIC を用いる場合には，必ず簡単なモデルから出発し，しだいに複雑性を増すモデルの階層構造を考える必要がある [5] が，状態を増やす場合でも，状態遷移を増やす場合でもこの条件を満たしていることがわかる

5 考察

5.1 提案するアルゴリズムを用いた結果

図 7 をみると，訓練用データに対する尤度は，状態数を 15 に固定したものがもっとも良く，構造を探索したモデルがそれに次いでいる．一方，認識用データに対して

5.2 尤度と認識率

以上の実験を通じて，情報量基準を最大化するモデルを探索することによって，その構造をあらかじめ固定する場合よりもデータをより良く表現できることが示せた．これは情報量基準の目的そのものであり，このようにして音素モデルを構成することにより，認識率が良くなる

ことが期待できる．ただし，表 3 に示された通り，このモデルは，あらかじめ状態数を 10 に固定した場合と認識率はほとんど変わらない．図 7 によれば，このモデルはデータをより良く表現している．にも関わらず，このような結果を得た原因は，Bayes の原理に基づいた認識方式にあると考えられる．これは，いくつかモデルがある中で，もっとも良い尤度を出したモデルを正しいものとするのであるが，このとき，そのモデルがある程度データを良く表しているならば，他との比較ではそのモデルが選ばれることになる．つまり，認識率のみを問題にするのならば，データをそれほど正確に表さなくてもよいであろう．

だからといって，適当にモデルを構成すればいいというわけではない．モデルを構成する場合，設計者が考えるべきことは，データを正確に表すことである．では，認識率を向上させるには，何をするべきなのであろう．ここで考えられるのは，カテゴリの問題である．例えば，ATR 研究用日本語音声データベースでのラベルとして用いられている音声記号は，37 である．では，全ての音素を 37 のカテゴリに分解するのが適切なのか，ということである．良く知られていることであるが，音素は前後にどのような音素が発声されるかによって影響を受ける．これを考慮に入れ，カテゴリをさらに分解した方が，各カテゴリの曖昧さが少なくなる．実際に CMU における自動音声認識システム SPHINX [9] では，前後の音素の並びから，triphone と呼ばれる 3 つの音素の並びを考え，当初 48 に分類されていた音素を最終的に 1076 個に細分化している．また，ATR の鷹見らの提案している SSS では [3]，一つの状態を持つ簡単なモデルから出発し，時間方向及びコンテキスト方向にモデルを複雑にしていくことで，HM-Net を構成し，音素認識の実験を行なっている．この場合，カテゴリの細分化も時間方向の分割と共に扱っている．

6　まとめ

5.2 節で説明したように，データを正確に表すということと，認識率とは必ずしも一致しないと考えられる．今回提案したアルゴリズムは与えられたカテゴリに属するデータを正確に表すことを目的としているから，認識率がそれほど向上していないことも理解できる．ただし，適切なカテゴリ分けがされている上で，それに対するモデルを構成するという問題に対して，このアルゴリズムを用いてモデルを構成することは有用である．

さらに認識率を向上させるためには，コンテキスト情報を用いて，カテゴリの細分化を行なう必要があるだろう．一方，カテゴリの細分化を行なう場合，カテゴリを分割する毎に各カテゴリのデータの数が少なくなることが問題となる．少ないデータからは，パラメタ数の多いモデルは推定できない．どのくらいのカテゴリに分解し，各カテゴリのデータ数をいくつにするかということに対しても，何らかの基準量を用いて，適切なものが選べれるべきである．

今後の課題としては，まず，適切なカテゴリ分けを行ない，それぞれのカテゴリに対して提案したアルゴリズムを用いることにより音素モデルを構成し，高い認識率で音素を認識するシステムの構成をすることが考えられる．それには，HMM の出力確率分布として，ガウス分布など，離散分布以外のものを用いることも必要であると考える．また，不特定和者に対するシステムの構成，音素を結合してできる単語のモデルの構成を行なうことも考えている．

参考文献

[1] 中川：“確率モデルによる音声認識”, 電子情報通信学会 (1988).

[2] 池田：“隠れマルコフモデルの生成に関する研究”, 修士論文, 東京大学, 計数工学科 (1993).

[3] 鷹見, 嵯峨山：“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モデル網の自動生成”, 技術研究報告, 電子情報通信学会誌 (1991).

[4] 赤池：“情報量規準 AIC とは何か–その意味と将来への展望”, 数理科学, 153, pp. 5–11 (1976).

[5] 竹内: “AIC 基準による統計的モデル選択をめぐって”, 計測と制御, **22**, 5, pp. 445–453 (1983).

[6] 武田, 匂坂, 片桐, 桑原：“研究用日本語音声データベースの構築”, 日本音響学会誌, **44**, 10, pp. 747–754 (1988).

[7] 武田, 匂坂, 片桐, 阿部, 桑原：“研究用日本語音声データベース利用解説書”, (株)ATR 自動翻訳電話研究所 (1986).

[8] Y. Linde, A. Buzo and R. M. Gray: “An algorithm for vector quantizer design”, IEEE Tr. Communications, **COM-28**, 1, pp. 84–95 (1980).

[9] K.-F. Lee: “Automatic Speech Recognition — The Development of the SPHINX System”, Kluwer Academic Publishers, Norwell, Massachusetts (1989).