

HMMの構造探索による音素モデルの生成

正員 池田 思朗[†]

Construction of Phone HMM Using Model Search Method

Shiro IKEDA, *Member*[†]

あらまし 本論文では、HMMの構造決定のためのアルゴリズムを提案する。音声認識で用いられるHMMの構造は、多くの場合知識や経験に基づき決定されている。しかしながら、この問題は確率モデルの構造決定問題として扱うことができる。すなわち、赤池情報量規準(AIC)などの尺度でモデルを評価し、最も良いモデルを選択するという方法で定式化できる。簡単な確率モデルにおいては、あらかじめ複数のモデルを用意し、最尤法でパラメタを推定した後にAIC等で最適なモデルを選択すれば良い。一方、HMMのようにパラメタ推定のための計算量が多い場合、効率良く、少ないモデルに対してのパラメタ推定のみで済ませたい。そのために、HMMの状態数、あるいは状態遷移を一つずつ増やしていき、そのつどパラメタ推定を行ない、AICでモデルを評価する方法をとり、これをモデル探索と呼ぶことにする。本アルゴリズムを用いてモデルを構成し、計算機上データによる実験、さらに音声データを用いて認識実験を行ない、データの統計的性質を反映したモデルがどのように構成されたかを調べ、本アルゴリズムを用いる効果を示す。

キーワード HMM, 音素モデル, モデル選択, 赤池情報量規準

1 まえがき

音声認識システムにおいてHMMを用いる際、一つの大きな問題となるのは、状態数やその間の状態遷移によって定義されるHMMの構造として、どのようなものを用いるかということである。実際のシステムを見てみると図1に示したBakisモデルを代表とする良く用いられる構造がある^{(9),(5)}。それぞれのシステムで状態数や出力確率分布などに差はあるものの、全ての音素に対しこのような構造のHMMを用いてシステムを構成する場合が多い。

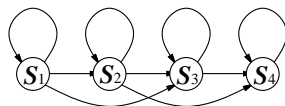


図1 Bakisモデルの例

一方、HMMの構造を決めるとは確率モデルの構造をいかに決めるかという問題であり、統計の分野でモデル選択と呼ばれる問題と等しい。モデル選択ではAIC(赤池情報量規準)を始めとするいくつかの情報量規準が提案されており、それらの規準を用いてモデルの良さを測り、適切なモデルを選択するのが一般的な手続きである。このようなモデル選択を用い、各音素に対しそれぞれの統計的性質を反映

した適切なモデルを選択すれば、全体のシステムとしてはより良いものとなるはずである。

通常モデル選択では複数のモデルのパラメタ推定を行なった後、AICによってモデルを選択する。そのときには無駄なモデルのパラメタ推定も行なうことになる。しかし、HMMに対してそのような通常モデル選択の方法を用いるのは問題がある。HMMのパラメタ推定では最尤推定を直接行なうことができず、Baum-Welchアルゴリズムを用い、繰り返し演算から推定するために時間がかかる。このようにパラメタ推定の計算量が多い場合、通常モデル選択の方法では効率が悪い。

本論文ではモデルの構造を段階的に変化させ、適切なモデルを“探索する”アルゴリズムを提案する⁽³⁾。具体的には、状態数あるいは状態遷移を増やしていき、そのつどパラメタの推定を行ないAICで評価する。これを繰り返し行ない、AICの規準で最も良いモデルを選ぶ。各音素毎に独立に本アルゴリズムを適用し、適切な構造のモデルを構成する。

このようにHMMの構造を変化させ、適切なモデルを選択しようという試みは他にも行なわれている^{(11),(4),(6)}が、問題設定やそのアルゴリズムに差が見られる。この点については5.2で述べる。

提案するアルゴリズムを用いて二つの実験を行なった。一つは、計算機内に定義した複数のHMMを信号源としてデータを生成し、そのデータを認識

[†]東京大学 工学部 計数工学科, 東京

Faculty of Engineering, University of Tokyo, 113,
Japan

する数値実験である．もう一つは ATR 音声研究用データベースを用いて /N/, /b/, /d/, /g/, /m/, /n/ 6 子音の音素認識実験である．

2つの実験を通して，本アルゴリズムによって得られたモデルが，データの統計的性質を反映した適切なモデルであることを示す．

2 AIC

2.1 AIC の定義

AIC^{(1),(14)} は，サンプルデータ $\{y_i\}$ ($i = 1, \dots, n$) に対する確率モデル $P(y_i|\theta)$ の良さを表す情報量規準である．(1) 式に示す通り，AIC は 2 つの項からなっており，これらの和によって適切なモデルを選択する規準を与えようというものである．AIC の第 1 項はデータに対するモデルの当てはまりの良さを表す対数ゆう度の項，第 2 項はモデルの複雑さを表すパラメタ数の項である．

$$\text{AIC} = (-2) \sum_{i=1}^n l(y_i|\theta^*) + 2p \quad (1)$$

ここで p はモデルの全パラメタ数， $l(y_i|\theta)$ は対数ゆう度，すなわち $l(y_i|\theta) = \ln P(y_i|\theta)$ ， θ^* は最ゆう推定を用いて求めたパラメタである．

一般にパラメタが多くなればモデルの表現力が増し，訓練用データに対するゆう度は良くなる．一方，パラメタの推定精度は落ち，同じカテゴリに属する新しいデータに対してのゆう度は悪くなる可能性が増す．このような状況で適切なモデルを選択するための一つの規準が AIC である．AIC の第 1 項は，ゆう度が良くなると小さくなる．一方，モデルの大きさに応じて第 2 項は大きくなり，ペナルティとして働く．従って AIC を小さくするモデルを選択することで，2 つの項のバランスのとれた適切な大きさのモデルが選択できると考えられる．

2.2 HMM の定式化及び HMM に対する AIC

(1) 式の AIC は最ゆう推定点で，

$$\left. \frac{\partial l(y|\theta)}{\partial \theta_i} \right|_{\theta^*} = 0 \quad \text{for } \forall i \quad (2)$$

が成り立つとして定義されている．一方 HMM のパラメタには最ゆう推定点で (2) 式を満たさないものが含まれている．そのため，このままでは AIC を

定義できない．以下では本論文で用いる HMM の定式化を行ない，HMM に対する AIC を定義する．

まず，本論文で用いる HMM の定式化をする．図 2 に例を示す．図中で I および F で示された状態ではデータは観測しないとする．従って図の HMM は 3 状態を持つことになる．出力確率分布は各状態に定義し，本論文で行なった実験のうち，人工データを用いたものでは離散分布，音素認識実験では連続分布 (多次元正規分布，但しそれぞれの成分は独立) であるとした．

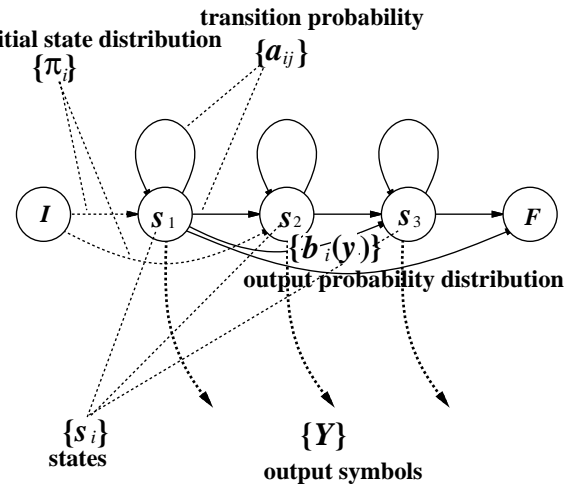


図 2 HMM の定義

状態 i の初期分布確率 $\{\pi_i\}$ ，状態 i から状態 j への状態遷移確率 $\{a_{ij}\}$ は，

$$\sum_i \pi_i = 1 \quad (3)$$

$$\sum_j a_{ij} = 1 \quad (4)$$

という拘束条件を持つ．

状態 i に定義される出力確率分布の密度関数 $b_i(y)$ は，離散分布とする場合，出力シンボルを o_k ($k = 1, \dots, K$) として次のように定義した．

$$b_i(y = o_k) = b_{ik} \quad (5)$$

$$\sum_k b_{ik} = 1 \quad (6)$$

拘束条件は (6) 式である．

また，連続分布の場合には y をベクトルの時系列と見なし， $b_i(y)$ を各次元が独立な多次元正規分布として定義した．この場合，離散分布のような拘束条件はなく，その密度関数は次元数を m ，ある時刻

のデータ y の成分を y_k として (7) 式の通りである .

$$b_i(y) = \left(\prod_{k=1}^m \frac{1}{2\pi\sigma_{ik}^2} \right)^{1/2} \exp \left\{ - \sum_{k=1}^m \frac{(y_k - \mu_{ik})^2}{2\sigma_{ik}^2} \right\} \quad (7)$$

本節の最初で述べたが, 状態遷移確率 $\{a_{ij}\}$ など拘束条件のある場合 (2) 式が成り立たない . このことを以下で示す .

HMM のパラメタ推定では $\{a_{ij}\}$ は (4) 式の拘束の下でゆわ度を最大にするように推定される . 最ゆう推定点では (8) 式が成り立っている (8) .

$$\begin{aligned} a_{ij} &= \frac{a_{ij} \frac{\partial P(y|\theta)}{\partial a_{ij}}}{\sum_{j'} a_{ij'} \frac{\partial P(y|\theta)}{\partial a_{ij'}}} = \frac{a_{ij} \frac{1}{P} \frac{\partial P(y|\theta)}{\partial a_{ij}}}{\sum_{j'} a_{ij'} \frac{1}{P} \frac{\partial P(y|\theta)}{\partial a_{ij'}}} \\ &= \frac{a_{ij} \frac{\partial l(y|\theta)}{\partial a_{ij}}}{\sum_{j'} a_{ij'} \frac{\partial l(y|\theta)}{\partial a_{ij'}}} \quad (8) \end{aligned}$$

(8) 式より $\partial l(y|\theta)/\partial a_{ij}$ が 0 でないことがわかる . 従って (2) 式は成り立たない . さらに状態 i からの各状態遷移 $a_{ij'}$ について (2) 式を求め, 両辺を $a_{ij'}$ で割ることにより, 各状態 i で次の関係が得られる .

$$\left. \frac{\partial l(y|\theta)}{\partial a_{ij}} \right|_{\theta^*} = \left. \frac{\partial l(y|\theta)}{\partial a_{ij'}} \right|_{\theta^*} \neq 0 \quad \text{for } \forall j' \text{ s.t. } a_{ij'} \neq 0 \quad (9)$$

但し, この値は状態毎に異なる .

各状態の拘束条件を用い, 状態毎に 0 でない状態遷移確率の任意の一つを, $a_{ik} = 1 - \sum_{j \neq k} a_{ij}$ として消去する . このとき, 残りの a_{ij} には形式的に拘束条件がなくなり (2) 式が成り立つことがわかる . このようにして拘束条件の分だけパラメタを減らしたモデルに対しては AIC を定義できる . 以下に本論文の実験で用いた 2 種類の HMM に対する AIC を定義しておく .

本論文で用いた HMM では, いずれの場合でも初期分布確率には (3) 式より 1 つ, 状態遷移確率に対しては (4) 式より状態と同じ数の拘束条件がある . 出力確率分布が離散分布の場合, 各状態に出力確率分布が定義されていることから, (6) 式の拘束条件は状態の数だけある . 従って, 全部で拘束条件は $(2 \times \text{状態数} + 1)$ 個あり, 各拘束条件で一つづつパラメタを消去できることから, 全体で $(2 \times \text{状態数} + 1)$

個のパラメタを消去することになる . AIC は全パラメタ数を p , 状態数を s として次のようになる .

$$\text{AIC} = (-2) \sum_{i=1}^n l(y_i|\theta^*) + 2\{p - (2s + 1)\} \quad (10)$$

出力確率分布を連続分布とした場合, 出力確率分布に関しては拘束条件がない . 初期分布確率と状態遷移確率の拘束条件は $(\text{状態数} + 1)$ 個であるから全パラメタ数を p , 状態数を s として AIC は次の通りである .

$$\text{AIC} = (-2) \sum_{i=1}^n l(y_i|\theta^*) + 2\{p - (s + 1)\} \quad (11)$$

3 提案するアルゴリズム

3.1 アルゴリズムの概要

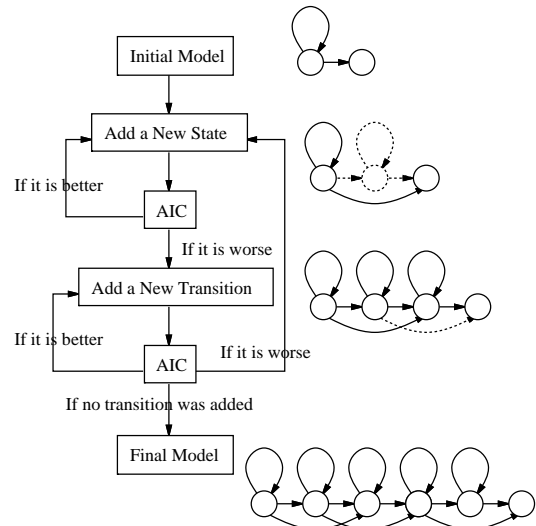


図3 アルゴリズムの流れ

図3に示したのが, 提案するアルゴリズムの流れ図である . HMM の構造は状態数とその間の状態遷移によって決まることから, これら一つづつ増やすことでモデルを探索する . まず, 状態を一つづつ増やしていくことで状態数を決める . このとき問題となるのは, 新しい状態をどこに定義するのが効果的か, 新しく定義した状態の初期値はどうするか, どこまで増やすか, の三つである . また, 状態数を決めた後, 新しく状態遷移を定義するが, このときも同様にいくつかの問題がある . 次節以降ではこれらの問題をどのように扱うかについて述べる .

3.2 状態数を増やす

3.2.1 被分割状態の決定

まず，現在ある状態を分割することで状態を増やしていく．どの状態を分割するか決定には，その状態の出力確率分布のエントロピー $Ent(i)$ に，その状態にどのくらいの時間留まるかの期待値 $ET(i)$ を乗じた量 $C(i)$ を用いて決定する．この $C(i)$ は HMM の出力確率分布がどのような形であっても定義できることから，本アルゴリズムは HMM の出力確率分布の形によらず適用できる． $C(i)$ によってエントロピーが大きく，留まる時間が長い状態を分割し，パラメタを増やすことでモデルの表現力を増やすことにする．出力確率分布のエントロピー $Ent(i)$ は，離散分布の場合は (12) 式，

$$Ent(i) = - \sum_{k=1}^K b_i(o_k) \ln b_i(o_k) = - \sum_{k=1}^K b_{ik} \ln b_{ik} \quad (12)$$

連続分布の場合は (13) 式となる．

$$Ent(i) = \sum_{k=1}^m \ln \sqrt{2\pi e \sigma_{ik}^2} \quad (13)$$

一方， $ET(i)$ は $\alpha_i(t)$ を前向き確率， $\beta_i(t)$ を後向き確率⁽⁹⁾として，(14) 式のように書ける．

$$ET(i) \propto \sum_{t=1}^{T-1} \sum_j \alpha_i(t) a_{ij} b_j(y_{t+1}) \beta_j(t+1) \quad (14)$$

これらを掛け合わせ，(15) 式の $C(i)$ を最大にする状態を分割する．

$$C(i) = \sum_{t=1}^{T-1} \sum_j \alpha_i(t) a_{ij} b_j(y_{t+1}) \beta_j(t+1) \times Ent(i) \quad (15)$$

3.2.2 初期値の設定

被分割状態を決めた後，どのように分割し，パラメタの初期値を決めるかが問題となる．

図4の2つのモデルを考える．この図は， S_1 を分割し， S_3 を新しく定義したことを示している．もし， $p'_2 = p_2$ ， $q_1 = p_1$ ， $q_2 = p_2$ ， $p'_1 + p'_3 = p_1$ であり， 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．そのようにして作ったモデルを初期値としてパラメタを推定し直せば，その結果得られるモデルのサンプルデータに対するゆう度は元のモデルよりも良くなる．

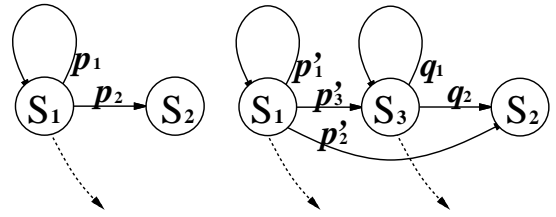


図4 状態分割

3.2.3 どこまで状態を増やすか

状態を増やし続けると，モデルが有限のサンプルデータに対して過学習をしてしまう．そこで，AIC を用いて状態数を増やすのを止めることにする．状態を増やしても AIC が減らない場合，それ以上は状態を増やさないことにする．

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 アルゴリズムによりパラメタを推定する．
4. AIC を計算し，前のモデルよりも良くなれば 1 へ戻る．そうでない場合は終了する．

3.3 状態遷移を増やす

3.3.1 状態遷移確率の変化

状態の数を決定しパラメタを推定した後，状態遷移も一つずつ増やしていく．まず，新しく定義する状態遷移の場所を決める．その後そこに小さな遷移確率を定義し，パラメタを推定する．これを繰り返していき，モデルを探索する．

新しく定義する状態遷移の場所は，加えることによってゆう度がどのくらい良くなるかを予測し，最も効果的な場所を選ぶことにする．予測には新しく定義する状態遷移の確率を小さな値 δ としたとき，B-W アルゴリズムの 1 ステップでの全パラメタの変化分 $(\hat{\theta} - \theta)$ と $\partial l(y|\theta)/\partial \theta$ という二つのベクトルの内積を用いる．

但し，この内積を計算する上では，全パラメタについて計算する必要はない．0 であった状態遷移 a_{ik} を新しく定義する際，他のパラメタ (a_{ik} 以外の状態遷移確率 $\{a_{ij}\}$ ，初期状態分布 $\{\pi_i\}$ ，出力確率分布 $\{b_i(y)\}$ のパラメタ) は最ゆう推定で求めたパラメ

タであることから，これらに対する $\partial l(y|\theta)/\partial\theta$ は 0 としてよい．従って $(\hat{\theta} - \theta)$ と $\partial l(y|\theta)/\partial\theta$ の内積を考える上では $(\hat{a}_{ik} - a_{ik})$ と $\partial l(y|\theta)/\partial a_{ik}$ の積のみが問題となるからである．この節ではまず B-W アルゴリズムの 1 ステップでの状態遷移確率の変化 $(\hat{a}_{ik} - a_{ik})$ について，次節では $\partial l(y|\theta)/\partial a_{ik}$ について見る．

まず， $\{a_{ij}\}$ に関する B-W アルゴリズムを示す^{(9),(8)}．本節以降では特に混乱のない限り $\partial_{ij} = \partial/\partial a_{ij}$ と表すことにする．

$$\hat{a}_{ij} = \frac{a_{ij}\partial_{ij}l(y|\theta)}{\sum_j a_{ij}\partial_{ij}l(y|\theta)} \quad (16)$$

B-W アルゴリズムでは， $a_{ij} = 0$ ならばその状態遷移確率は変化せず，HMM の構造は変化しない．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場合， \hat{a}_{ik} は次式のようになる．

$$\hat{a}_{ik} = \frac{\delta\partial_{ik}l(y|\theta)}{\sum_{j':j' \neq k} a_{ij'}\partial_{ij'}l(y|\theta) + \delta\partial_{ik}l(y|\theta)} \quad (17)$$

$\partial_{ik}l(y|\theta)$ は必ず有限値をとることと，(9) 式の結果及び(4) 式の拘束条件より，最ゆう推定点では 0 でない a_{ij} による $\partial_{ij}l(y|\theta)$ を用いて(18) 式が成り立つ．

$$\sum_{j'} a_{ij'}\partial_{ij'}l(y|\theta) = \left(\sum_{j'} a_{ij'} \right) \partial_{ij}l(y|\theta) = \partial_{ij}l(y|\theta) \quad (18)$$

この結果を用い， δ を十分小さいとして(17) 式を展開すると，

$$\begin{aligned} \hat{a}_{ik} &= \frac{\delta\partial_{ik}l(y|\theta)}{\partial_{ij}l(y|\theta) + \delta\partial_{ik}l(y|\theta)} \\ &\simeq \delta \frac{\partial_{ik}l(y|\theta)}{\partial_{ij}l(y|\theta)} \left[1 - \delta \frac{\partial_{ik}l(y|\theta)}{\partial_{ij}l(y|\theta)} \right]. \end{aligned} \quad (19)$$

(19) 式から， δ の 2 次以上の項を無視すると a_{ik} の増加分は，

$$\begin{aligned} \hat{a}_{ik} - a_{ik} &= \hat{a}_{ik} - \delta \\ &\simeq \delta \left[\frac{\partial_{ik}l(y|\theta)}{\partial_{ij}l(y|\theta)} - 1 \right] \\ &\quad j \neq k \wedge a_{ij} \neq 0 \end{aligned} \quad (20)$$

とかける．この値が正であれば，その状態遷移確率は増加すると考えらる．

3.3.2 ゆう度の変化

一方， $\partial_{ik}l(y|\theta) = \partial l(y|\theta)/\partial a_{ik}$ は(21) 式で与えられる^{(9),(8)}．

$$\partial_{ik}l(y|\theta) = \frac{\sum_{t=1}^{T-1} \alpha_i(t)b_k(y_{t+1})\beta_k(t+1)}{\sum_i \alpha_i(T)} \quad (21)$$

従って，新しく状態遷移 a_{ik} を定義したときの $(\hat{\theta} - \theta)$ と $\partial l(y|\theta)/\partial\theta$ の内積 $\Delta_{ik}(y, \theta)$ は，

$$\begin{aligned} \Delta_{ik}(y, \theta) &\equiv (\hat{\theta} - \theta) \cdot \frac{\partial l(y|\theta)}{\partial\theta} = (\hat{a}_{ik} - \delta) \cdot \partial_{ik}l(y|\theta) \\ &\simeq \delta \left[\frac{\partial_{ik}l(y|\theta)}{\partial_{ij}l(y|\theta)} - 1 \right] \partial_{ik}l(y|\theta) \end{aligned} \quad (22)$$

となる． $\partial_{ik}l(y|\theta)$ の値は(21) から求まる．このように定義された $\Delta_{ik}(y, \theta)$ を候補に上がっている状態遷移それぞれについて求め，最も大きい値を示した状態遷移を採用する．実際には各 $\Delta_{ik}(y, \theta)$ で δ は共通なので， δ の値は考える必要がない．また，以上の方法では，全ての $\Delta_{ik}(y, \theta)$ を求める際に一組の $\{\alpha_i(t)\}$ ， $\{\beta_i(t)\}$ のみが必要であることから B-W アルゴリズムの 1 ステップの計算を一回行なえばすむ．

採用した後はそこに小さな値を代入し，B-W アルゴリズムを行い，新しいモデルを作る．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定める．

4 実験

4.1 人工データを用いてのシミュレーション

まず，人工的に作った 5 つの HMM 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．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すなわち多項分布で，6 つのシンボルを出力するものとする．従って得られる時系列信号はこれらの 6 つのシンボルの列となる．5 つのモデルは，それぞれ状態数，状態遷移，出力確率分布のいずれかまたは全てで異なっており，これに応じてそれぞれが出力する信号も異なっている．それぞれの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モデルの構造及び状態遷移確率を図 5 に示す．

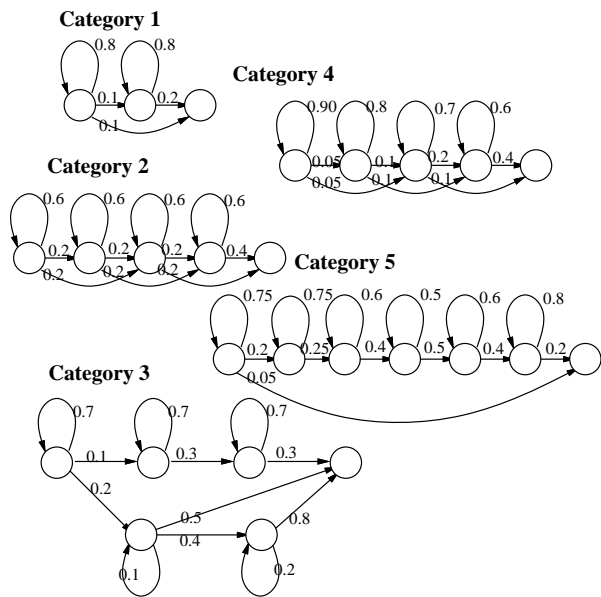


図5 信号発生用の HMM

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

アルゴリズムの結果、各カテゴリに対して得られた HMM の構造及び状態遷移確率を図6に示す。これらの HMM の出力確率分布も離散分布であり、出力シンボルの種類は6つである。この図から、信号発生用の HMM の構造を良く反映しているものもあるが、そうでないものもあることがわかる。

比較のために図7に示した3状態、5状態、7状態の HMM を用いて、同様の実験を行なった。本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。表には誤認識率を示した。

表1 認識率

信号発生用	3 状態	5 状態	7 状態	モデル探索
	92.56%	84.92%	90.22%	90.68%
				91.94%

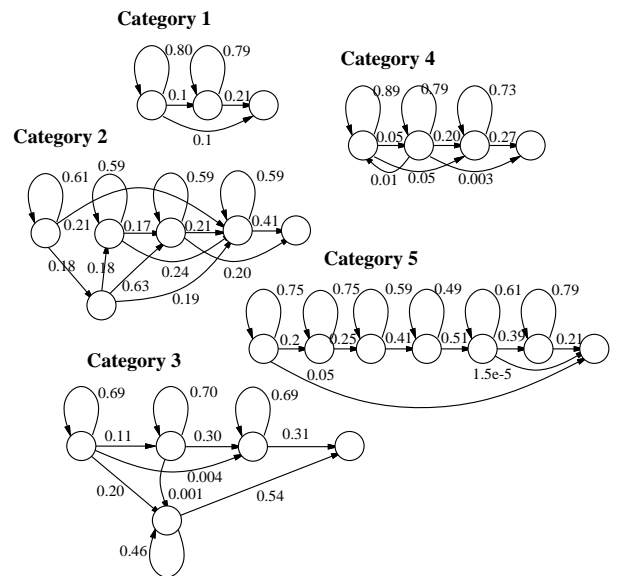


図6 構造を変化させた結果

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8に示す。これを見ると、AICを用いてモデル探索したモデルが、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。

4.2 音声データを用いての実験

提案するアルゴリズムを用いて、実際の音声データに対し HMM を生成し、認識を行なった。

使用したデータ ATR 研究用日本語音声データベース⁽¹³⁾の男性話者1名(MAU)。各音素データはラベルに従って切り出した⁽¹²⁾。

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.

訓練用データ 単語発声データ(5240単語)の偶数番。

認識実験用データ 単語発声データの奇数番。

認識タスク /N/, /b/, /d/, /g/, /m/, /n/ の6子音。HMMの出力確率分布 32次元の多次元正規分布,

各次元は独立。

データの形式 logpow, cep(15), Δlogpow, Δcep(15) からなる 32次元ベクトル。

状態分割の制限 1モデル当たりの状態数を最大20に制限。

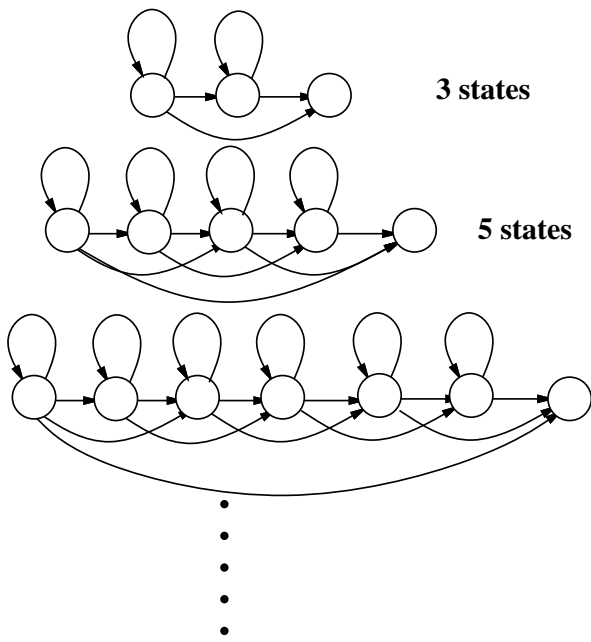


図7 比較のための HMM

アルゴリズムを用いて出来上がったそれぞれのモデルの状態数および状態遷移数を表2に、また/d/に対して得られた HMM の構造を図9に示す。

表2 モデル探索によって求めたモデルの構造

子音 (訓練用データ数)	N (503)	b (218)	d (202)	g (260)	m (471)	n (260)
状態数	20	20	20	20	20	20
状態遷移の数	83	68	67	105	78	73

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、4,8,12,16,20の状態数をもつ図7のようなHMM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。表からわかるように、提案するアルゴリズムを用いた場合、他のモデルを用いるよりもより良い認識率を上げている。

構造を変化させていく過程で、訓練用、認識用のデータに対してゆう度がどのように変わっていくかを図10に示した。図は/g/に対する各モデルのゆう度を示しており、比較のため、状態数が4,8,12,16,20であり、構造を探索せずに全ての音素に対して同じ構造のモデルを用いた場合の結果も図中に●、×で示した。他の音素に対してもおおよそ同じような傾向が現れている。

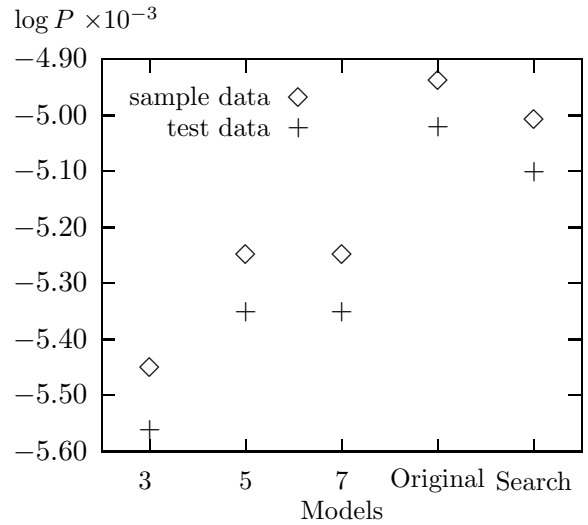


図8 用いたモデル間でのゆう度の比較

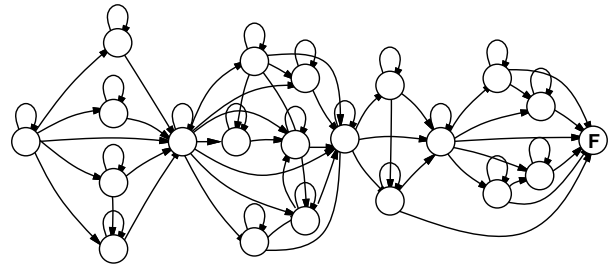


図9 /d/に対して得られた HMM の構造

5 考察

5.1 提案するアルゴリズムの効果

図10をみると、それぞれのモデルのゆう度は訓練用データ、認識用データ両方に対して提案するアルゴリズムを用いたものが最も良い。また、認識結果も本アルゴリズムを用いたものが最も良くなっている。これは構造探索を行なったことによって他のモデルよりもデータを良く表現するモデルを作成で

表3 誤認識率

モデルの種類	訓練用データ		認識用データ	
	第一位	第三位	第一位	第三位
比較用 4 状態	23.72%	2.66%	24.91%	2.38%
比較用 8 状態	16.30%	1.25%	17.12%	1.22%
比較用 12 状態	13.38%	0.47%	15.42%	0.95%
比較用 16 状態	13.06%	0.42%	15.37%	0.74%
比較用 20 状態	12.23%	0.57%	13.67%	1.11%
モデル探索	7.99%	0.05%	11.34%	0.74%

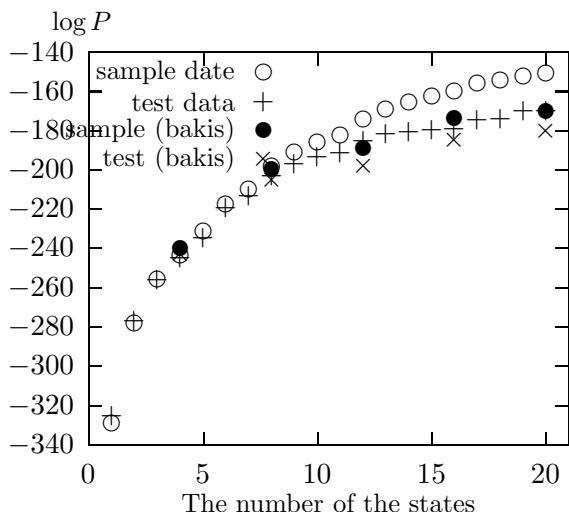


図10 用いたモデル間でのゆう度の比較

きたことを示している。

数値実験ではデータの統計的構造が単純であることから、状態数もさほど増えず、AICを用いて打ち切ることによって適切なモデルを探索できた。しかし、音素認識実験では、音声データの統計的な構造が複雑であり、またデータの数も多かったことから、AICを用いて状態数を打ち切る前に状態数の制限が効いてしまった。AICで打ち切られるまでモデルを大きくすると、非現実的な数の状態数を選択することになってしまう。

また図10から、状態数が同じ場合でも、モデルを探索した時の方がゆう度は良くなっている。これは何を意味しているのだろうか。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るという問題がある。提案するアルゴリズムで用いる初期モデルでは状態が一つしかなく、B-Wアルゴリズムで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、本アルゴリズムでは極小解になる可能性を減らせるのではないかという期待ができる。

このことを確かめるために、一つの実験を試みた。提案するアルゴリズムの結果求めたモデルの構造のみを用い、パラメタは状態遷移確率に関しては等確率に、出力確率分布に関しては平均を0、分散を1として、B-Wアルゴリズムで推定し直した。音素/g/に対するそれぞれのモデルのゆう度を図11に示す。図中の左の二本はアルゴリズムで探索の結果求めたモデル、右の二本がその構造のみを用いてパラメタを推定し直したものである。訓練用、認識用

のそれぞれのデータに対して、探索を通して求めたモデルの方が良い結果を示している。この傾向は他の音素に対しても同様であった。

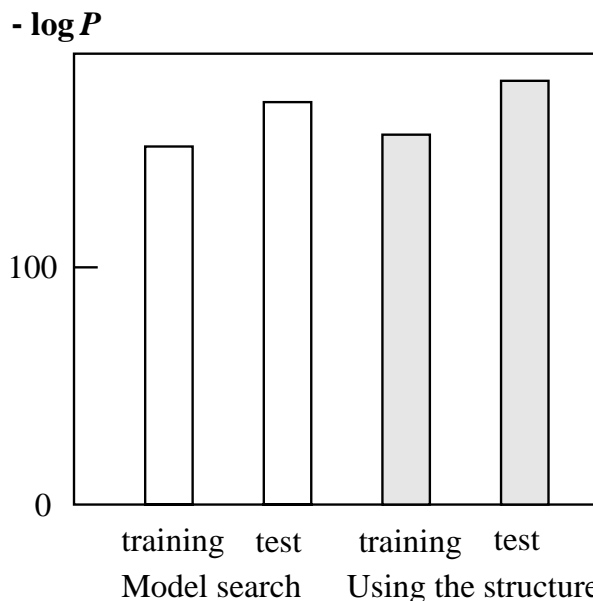


図11 /g/に対するゆう度

そうして求め直したモデルを用いて認識実験を行った。その結果を表4に示す。アルゴリズムを用いたものの方が良い結果を示していること分かる。

表4 誤認識率

モデルの種類	訓練用データ		認識用データ	
	第一位	第三位	第一位	第三位
構造のみ	8.57%	0.47%	12.45%	2.19%
モデル探索	7.99%	0.05%	11.34%	0.74%

5.2 他の手法との比較

本実験で行なった行なった実験の結果は、HMMを用いて行なわれた実験としてはそれほど良くない。子音の認識実験では様々なHMMでの比較実験が行なわれ^{(10),(2)}、特に子音部分のみでなく後続母音とのわりもデータとして用いた方がより認識率が上がることや、データとしてスペクトルの動的特徴パラメタを組み込んだ方が良いこと、また確率分布としても混合分布を用いた方が良いことなどが示されている。これとは別に、CMUにおける自動音声認識システム SPHINX⁽⁷⁾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. このようにデータのカテゴリを分解して各カテゴリの統計的なばらつきを小さくした方が認識率が良くなることも知られている.

但し, 本実験の目的は, 全ての音素に対し同じ構造の HMM を用いるよりも, それぞれの音素に適切なモデルをモデル探索により構成することで, より良いシステムとなることを示すことにあった. このことは本論文で行なった比較実験によって確かめられた. 今後の課題として, さらに認識率をあげるためには, 前に挙げた知見を利用しなければならない.

本論文で提案するアルゴリズム同様 HMM の構造を変化させていくというアプローチは他にもいくつか^{(11),(4),(6)} 行なわれている. ATR の鷹見らの提案した逐次状態分割法 (Successive State Splitting: SSS)⁽¹¹⁾ では, 一つの状態を持つモデルから出発し, 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隠れマルコフ網 (HMnet) を構成し, 音素認識の実験を行なっている. この場合, 前後の音素の系列から各音素モデルを細分化し, データのカテゴリを細分化している. また, 岩山らの方法⁽⁴⁾ では, あらかじめ大きなモデルを用意しておき, そこから状態間のリンクを削除していくことでモデルの構造を変化させている. Kamp の方法⁽⁶⁾ では, 複数の単語モデルがあるときに, 相互の関係から, 認識率を変化させないように状態を減らしていき, モデルの状態数を減らすことを提案している.

本論文で提案するアルゴリズムは, 第 1 に状態とその間の状態遷移を増やすことで適切なモデルを構成することが目的である. この点で岩山らのアルゴリズムや, Kamp のアルゴリズムとは異なっている. 第 2 に, 本論文の目的は各カテゴリに対し適切なモデルを独立に構成することであり, カテゴリを細分化することは考えていない. この点で鷹見らの方法, そして Kamp の方法とも異なっている. モデルを精度良く推定するためにはデータの量が必要であることから, カテゴリの数と各カテゴリのデータの量とのバランスをとっていかなければならない. この問題に関しても, 何らかの規準量を用いて適切なものを選ぶべきであり, これも今後の課題である.

6 むすび

HMM の構造を決定する方法として, その構造を変化させ, 探索するアルゴリズムを提案した.

このアルゴリズムを用いて数値実験, 及び音素認識実験を行なった. その結果全てのカテゴリに対して同じ構造のモデルを用いる場合よりも良い認識率が上げられることが確かめられ, 本アルゴリズムの有用性が示せた.

今後の課題としては, 基本的な HMM として, 様々な知見を生かし, より認識率の良いモデルを構成することを考えている. さらに, データに関してもわたりを含むデータを利用し, 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定義し, 音素認識システムの構成をすることが考えている. また, 不特定話者に対するシステムの構成, 文節, 単語モデルの構成も行なっていきたい.

謝辞 本研究を行なう機会を与えて下さった, 中野馨助教授に感謝します. また阪口豊助手, また大学院生の下平英寿氏にはさまざまな助言を頂きました. 感謝致します. なお, 本研究の実験では ATR 研究用日本語音声データベースを使用しました.

文 献

- (1) 赤池弘次. 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, No. 153, pp. 5–11, March 1976.
- (2) 花沢利行, 川端豪, 鹿野清宏. Hidden Markov モデルによる日本語有声破裂音の認識. 音響誌, Vol. 45, No. 10, pp. 776–785, October 1989.
- (3) 池田思朗. HMM の構造探索による音素モデルの生成. 信学技報, Vol. SP93-26, pp. 17–24, June 1993.
- (4) 岩山真, Nitin Indurkha, 元田浩. リンク削除に基づく隠れマルコフモデルの自動構成法. 人工知能研究会資料, Vol. 86-10, pp. 73–80, January 1993.
- (5) Frederick Jelinek. Continuous Speech Recognition by Statistical Methods. *Proc. IEEE*, Vol. 64, No. 4, pp. 532–556, April 1976.
- (6) Yves Kamp. State Reduction in Hidden Markov Chains Used for Speech Recognition. *IEEE Trans. Acoust., Speech & Signal Process.*, Vol. ASSP-33, No. 4, pp. 1138–1145, October 1985.
- (7) Kai-Fu Lee. *Automatic Speech Recognition — The*

Development of the SPHINX System. Kluwer Academic Publishers, Norwell, Massachusetts, 1989.

- (8) S.E. Levinson, L.R. Rabiner, and M.M. Sondhi. An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition. *Bell Syst. Tech. J.*, Vol. 62, No. 4, pp. 1035–1073, April 1983.
- (9) 中川聖一. 確率モデルによる音声認識. 電子情報通信学会, 1988.
- (10) 中川聖一, 平田好充, 橋本泰秀. 連続出力分布型 HMM による日本語音韻認識. 音響誌, Vol. 46, No. 6, pp. 486–496, June 1990.
- (11) 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モデル網の自動生成. 信学技報, Vol. SP91-88, pp. 57–64, December 1991.
- (12) 武田一哉, 匂坂芳典, 片桐滋, 阿部匡伸, 桑原尚夫. 研究用日本語音声データベース利用解説書. ATR 自動翻訳電話研究所, 1986.
- (13) 武田一哉, 匂坂芳典, 片桐滋, 桑原尚夫. 研究用日本語音声データベースの構築. 日本音響学会誌, Vol. 44, No. 10, pp. 747–754, 1988.
- (14) 竹内啓. AIC 基準による統計的モデル選択をめぐって. 計測と制御, Vol. 22, No. 5, pp. 445–453, March 1983.



池田 思朗

平3 東大・工・計数卒, 平5 同大学院・工・計数修士了. 現在, 同博士課程在学中. 音声認識, 統計的モデル選択の研究に従事.