音声言語情報処理 28－5
（1999．10．29）

平滑化部分隠れマルコフモデルによる音声認識

古山 純子　　小林 哲則
早稲田大学 理工学部
〒169-0072 東京都新宿区大久保3-4-1
早稲田大学理工学部電気電子情報工学科 小林哲則研究室
E-mail:{furuyama,koba}@tk.elec.waseda.ac.jp

あらまし： 我々はHMMの欠点を補う新たな確率モデルとして、部分隠れマルコフモデル(PHMM)を提案してきた。PHMMは音声認識やジェスチャ認識に有効であることを示してきたが、このモデルはHMMに比べ高次の統計量を扱っているため、学習データ量が十分でない時、モデルの信頼性が低下してしまう可能性がある。そこで本研究では、PHMMの高次元の確率をHMMの低次の統計量で平滑化した、平滑化部分隠れマルコフモデルを提案した。
このモデルを単語音声認識実験に適用したところ、出力確率を平滑化することで、差分無しの特徴量を用いた場合で50.0%、差分有りの特徴量を用いた場合で75.0%、PHMMの誤認識率を改善することができた。これは、HMMの誤認識率を82.5%(差分無し)および85.7%(差分有り)改善する結果となった。

Speech Recognition using Smoothed Partly-Hidden Markov Model

Furuyama Junko　　Tetsunori Kobayashi
School of Science and Engineering ,Waseda University
3-4-1 Okubo, Shinjyuku-ku, Tokyo 169-0072 JAPAN
E-mail:{furuyama,koba}@tk.elec.waseda.ac.jp

Abstract : We solved HMM's shortcomings by introducing the modified second order Markov Model, Partly-Hidden Markov Model (PHMM), in previous work. It was shown that the model is effective for speech and gesture recognition, however, the model becomes unreliabile without sufficient training data, since the model uses hight order statistics than HMM. In this paper, we propose Smoothed Partly-Hidden Markov Model (SPHMM), in which the statistics of PHMM is smoothed with lower order statistics.
Word recognition test using SPHMM shows that the error rate was reduced by 50.0% (without delta) and 75.0% (with delta) compared with PHMM, and 82.5% (without delta) and 85.7% (with delta) compared with HMM.

1 はじめに

音響モデルは、音声認識の根幹をなす重要な技術であり、その精密化に向けて多くの研究が行なわれている [1][2][3][4] [5]。我々は、2重のマルコフモデルから発して、一方を隠れ状態に、他方を可観測な状態におくことで、単純な HMM より過渡部の表現能力に優れる新たなモデル (PHMM) を提案し、このモデルが音声認識をはじめジェスチャ認識など、時系列パターン認識に有効であることを示してきた [6][8]。PHMM は、モデルの精密化に伴い尤度の信頼性が向上するとともに、現象の文脈依存性をモデル内で表現できるなどの特徴を持っている。しかし、このモデルは HMM に比べ高次の統計量を扱うため、学習データが少ない時、モデルの信頼性が失われる可能性がある。そこで、本研究では、PHMM の欠点を補うために、PHMM の遷移確率、出力確率それぞれを、低次の統計量を用いて平滑化した、平滑化部分隠れマルコフモデルを提案する。平滑化の効果により、PHMM の持つ精密性と、HMM の耐性を合わせ持つ確率モデルを構成することを目指す。

本論文では、2 章において部分隠れマルコフモデルについて概観した後、3 章において平滑化部分隠れマルコフモデルを導入する。4 章において、平滑化部分隠れマルコフモデルの評価実験について述べる。

2 部分隠れマルコフモデル

2.1 PHMM の概要

時刻 t における出力ベクトル x_t が過去 K 個の出力 $x_{t-K},\cdots,x_{t-1}$ の条件付確率によって与えられる確率過程を、次のような 2 重マルコフモデルで表現する。

$$P_r(x_t|x_{t-K}x_{t-K+1}\cdots x_{t-1}) = P_r(x_t|S_i^f S_j^s) \quad (1)$$

ここで状態 S_i^f(F 状態) は $x_{t-K}x_{t-K+1}\cdots x_{t-2}$ の出力列に対応し、状態 S_j^s (S 状態) は出力 x_{t-1} に対応して与える。仮にこれらの写像が共に 1 対 1 対応であればモデルはマルコフモデルと等価であり、共に写像が確率的であれば HMM と等価である。

本モデルでは、出力列 $x_{t-K}\ x_{t-K+1}\cdots x_{t-2}$ から F 状態 S_i^f への写像は確率的に行ない、出力 x_{t-1} から S 状態 S_j^s への写像は一意に決める。これにより、条件部の半分が多くの出力列で共有されることから、F 状態 S_i^f の数を抑えることができ、結果としてモデルの複雑さも抑えることができる。また、x_t の出力確率は、F 状態 S_j^f だけでなく、S 状態 S_j^s すなわち x_{t-1} によっても条件付けられるため、区分定常以上の複雑な過程を扱うことが可能になる。

このモデルを 部分隠れマルコフモデル (Partly-Hidden Markov Model : PHMM) と定義する。

2.2 確率計算とパラメータ

PHMM を用いて出力列 $x_1x_2\cdots x_T$ が観測される確率 $P_r(x_1x_2\cdots x_T)$ を考える。

時刻 t における F 状態を s_t^f、時刻 t における S 状態を s_t^s とする。時刻 $t+1$ における S 状態は、定義により一意に

$$s_{t+1}^s = x_t \quad (2)$$

と決まる。また、時刻 $t+1$ における F 状態 s_{t+1}^f は、s_t^f だけでなく、s_t^f と s_t^s の条件付確率として与えられるものとする。

このとき、出力列 $x_1x_2\cdots x_T$ が、F 状態遷移 $s_1^f s_2^f\cdots s_T^f$ S 状態遷移 $s_1^s s_2^s\cdots s_T^s$ から生起する確率を P_s とおけばその値は、前述までの議論より次式で表わされる。

$$\begin{aligned} P_s &= P_r(x_1x_2\cdots x_T, s_1^f s_2^f\cdots s_T^f, s_1^s s_2^s\cdots s_T^s) \\ &\simeq P_r(s_1^f, x_0)P_r(x_1|s_1^f x_0) \\ &\quad \times \prod_{t=1}^{T-1} P_r(s_{t+1}^f|s_t^f x_{t-1})P_r(x_{t+1}|s_{t+1}^f x_t) \end{aligned} \quad (3)$$

PHMM における状態と出力の関係を図 1 に示す。通常の HMM やセグメント HMM、Δ パラメータを含むモデルが数フレームの出力列あるいはそれと等価な情報の同時出力確率を扱うのに対して、このモデルは、過去の出力に関する条件付きの出力確率を扱い、さらに出力確率のみならず、状態遷移も過去の出力に依存している点が大きくことなる。

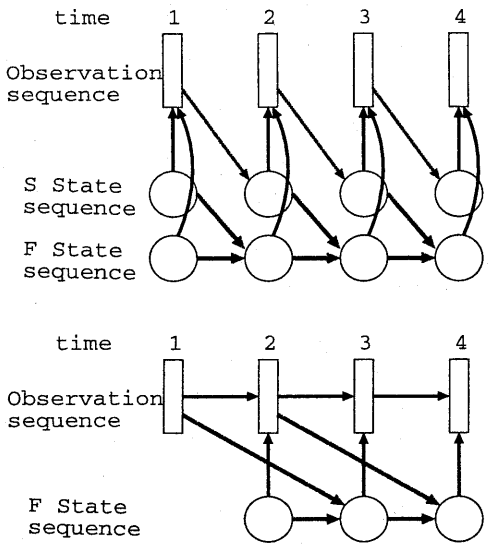

図 1: PHMM における状態出力関係図 (上段:S 状態を表示、下段:S 状態を $s^s_{t+1} = x_t$ として非表示)

また、ベイズの定理より

$$P_r(s^f_{t+1}|s^f_t x_{t-1}) = \frac{P_r(s^f_{t+1}|s^f_t)P_r(x_{t-1}|s^f_t s^f_{t+1})}{P_r(x_{t-1}|s^f_t)}$$

$$P_r(x_{t+1}|s^f_{t+1}x_t) = \frac{P_r(x_t x_{t+1}|s^f_{t+1})}{P_r(x_t|s^f_{t+1})}$$

であるから、式 (3) は、

$$P_s \simeq P_r(s^f_1)P_r(x_0x_1|s^f_1) \times \prod_{t=1}^{T-1} \frac{P_r(s^f_{t+1}|s^f_t)P_r(x_{t-1}|s^f_t s^f_{t+1})}{P_r(x_{t-1}|s^f_t)} \cdot \frac{P_r(x_t x_{t+1}|s^f_{t+1})}{P_r(x_t|s^f_{t+1})} \quad (4)$$

となる。

求める $P_r(x_1x_2\cdots x_T)$ は可能性のあるすべての F 状態遷移 $s^f_1 s^f_2 \cdots s^f_T$ の組合せに対し、式 (4) を加えればよい。

以上の議論から、PHMM を決めるパラメータ $\boldsymbol{\theta}$ は、次の 5 種類となる。

- $\pi_i = P_r(s^f_1 = S^f_i)$
  初期時刻に F 状態が S^f_i にある確率
- $a_{ij} = P_r(s^f_{t+1} = S^f_j | s^f_t = S^f_i)$
  時刻 t に F 状態が S^f_i にあり、次時刻に S^f_j に遷移する確率
- $b_i(x_{t-1}) = P_r(x_{t-1}|s^f_t = S^f_i)$
  時刻 t に F 状態が S^f_i で、前時刻出力がが x_{t-1} である確率
- $c_{ij}(x_{t-1}) = P_r(x_{t-1}|s^f_t = S^f_i s^f_{t+1} = S^f_j)$
  F 状態が時刻 t に S^f_i から S^f_j に遷移するときに、前時刻の出力が x_{t-1} である確率
- $d_j(x_{t-1}x_t) = P_r(x_t x_{t-1}|s^f_t = S^f_j)$
  F 状態が S^f_j であるときに前時刻の出力が x_{t-1} で現時刻の出力が x_t である確率

2.3 モデルの自由度

left-to-right 型の HMM と PHMM について、モデルの自由度 (パラメータ数) を比較する。特徴量の次元数を D とし、各パラメータの確率分布を単一の正規分布で表現するときの一状態当たりパラメータの数は、表 1 の様になる。 最大次数 D^2 の係数は、HMM で 1/2 であり、PHMM では 3 であるので、大雑把にいって、PHMM は HMM に比べ 6 倍程度 自由度が大きいことになる。

表 1: 1 状態当たりの自由度の比較

	HMM	PHMM
自由度	$\frac{1}{2}D^2 + \frac{3}{2}D$	$3D^2 + 6D$

3 平滑化部分隠れマルコフモデル

3.1 SPHMM の概要

PHMM は HMM より高次元の統計量を扱っている。そのため、学習データ量が十分でない時には、モデルの信頼性が失われている可能性がある。そのため、PHMM の高次元の統計量を HMM の低次の統計量で平滑化することが有効だと考えられる。

PHMM の遷移確率、出力確率それぞれを HMM の各確率の相乗平均で置き換えると、確率計算式 (4) は以下のようになる。

$$P_s \simeq P_r(s^f_1)P_r(x_0x_1|s^f_1) \times \prod_{t=1}^{T-1} \left\{ \left[\left(\frac{P_r(s^f_{t+1}|s^f_t)P_r(x_{t-1}|s^f_t s^f_{t+1})}{P_r(x_{t-1}|s^f_t)} \right)^{w_t} \cdot Pr(s^f_{t+1}|s^f_t)^{(1-w_t)} \right] \cdot \left[\left(\frac{P_r(x_t x_{t+1}|s^f_{t+1})}{P_r(x_t|s^f_{t+1})} \right)^{w_o} \cdot Pr(x_{t+1}|s^f_{t+1})^{(1-w_o)} \right] \right\} \quad (5)$$

ここで、w_t は遷移確率の平滑化の重みを表し、w_o は出力確率の平滑化の重みを表している。相乗

平均を用いて平滑化しているため、可能性のある全てのF状態遷移の組合せに対し式(5)を加えても確率が1にはならないが、確率計算において対数をとって加算している関係上相加平均ではなく相乗平均を用いた。

このモデルを、平滑化部分隠れマルコフモデル (Smoothed Partly-Hidden Markov Model : SPHMM) とする。

3.2 SPHMMにおけるViterbiアルゴリズム

先に述べたパラメータを利用することにより、SPHMMにおいて高速な $Pr(x_1 x_2 \cdots x_T)$ 求め方を考える。

ここでは、F状態が隠れ状態にあたるから、SPHMMのViterbiアルゴリズムは、$x_0 x_1 \cdots x_T$ の出力確率を最大にするF状態の遷移と、その最大確率値を求めることにあたる。このときの最適なF状態遷移列を $\hat{\boldsymbol{S}} = \hat{s}_1^f \hat{s}_2^f \cdots \hat{s}_T^f$、$\hat{\boldsymbol{S}}$ が与える $x_1 x_2 \cdots x_T$ の出力を P_v とする。

時刻 t にF状態が S_j^f に至る状態遷移のうち、x_0 から x_t までを出力する確率の最大値をあたえるものを考え、そのときの確率を $\delta(j,t)$、1時刻前のF状態へのポインタを $\psi(j,t)$ とすれば、最適状態系列を求める全体的な手法は、次のように書くことが出来る。

1) 初期化 $(t = 1, 1 \leq j \leq N)$

$$\begin{aligned} \delta(j,1) &= \pi_j d_j(x_0, x_1) \\ \psi(j,t) &= 1 \end{aligned} \tag{6}$$

2) 繰り返し $(2 \leq t \leq T, 1 \leq i \leq N)$

$$\begin{aligned} \delta(j,t) = \max_i \Bigg[& \delta(i,t-1) \\ & \cdot \left(\frac{a_{ij} c_{ij}(x_{t-2})}{b_i(x_{t-2})} \right)^{w_t} \cdot {a_{ij}}^{(1-w_t)} \Bigg] \\ & \cdot \left(\frac{d_j(x_{t-1}, x_t)}{b_j(x_{t-1})} \right)^{w_o} \cdot b_j(x_t)^{(1-w_o)} \end{aligned} \tag{7}$$

$$\begin{aligned} \psi(j,t) = \arg\max_i \Bigg[& \delta(i,t-1) \\ & \cdot \left(\frac{a_{ij} c_{ij}(x_{t-2})}{b_i(x_{t-2})} \right)^{w_t} \cdot {a_{ij}}^{(1-w_t)} \Bigg] \\ & \cdot \left(\frac{d_j(x_{t-1}, x_t)}{b_j(x_{t-1})} \right)^{w_o} \cdot b_j(x_t)^{(1-w_o)} \end{aligned} \tag{8}$$

3) 終了 (t=T)

$$\begin{aligned} P_v &= \delta(J,T) \\ \hat{s}_t^f &= S_J^f \end{aligned} \tag{9}$$

4) パス（状態遷移）バックトラック $(t = T-1, T-2, \cdots, 1)$

$$\hat{s}_t^f = S^f_{\psi(f(\hat{s}_{t+1}^f), t+1)} \tag{10}$$

ただし、N はF状態の数であり、$f()$ はF状態を引数にとってその添字の番号を返す関数である。

$\hat{\boldsymbol{S}}$ は、与えられたモデルパラメータによる最適なセグメンテーションの結果を表す。

3.3 SPHMMの学習

SPHMMの学習にはセグメンタルk平均アルゴリズムを用いた。即ち、平滑化の重み w_t、w_o の各値毎に、モデルパラメタの第i回近似解を用いてセグメンテーションを行ない、そのセグメントを用いて第i+1回近似解を求める処理を繰り返す。

4 単語音声認識実験

HMM、PHMM、SPHMMの性能を比較するため単語音声認識実験を行なった。

4.1 実験条件

(a) 実験データ

ATRの音素連鎖バランス単語216単語を20人の被験者に対し、1語につき各5回話したものをデータとして収集した。そのうち1回分を認識データ、残りの4回分を学習データとし、認識に使うデータを5種類変化させて実験を行なった。

(b) 特徴量

次の2種類の差分無し特徴量 z_t-13Dと、差分有り特徴量 z_t-26Dを用いた。

z_t-13D = (MFCC, パワー)

z_t-26D = (z_t-13D, ΔMFCC, Δパワー)

(c) 確率モデルの構造

PHMM、HMM ともに確率分布関数は単一正規分布とした。PHMM の F 状態数及び HMM の状態数は、各単語毎にそれを構成する音素数として与え、構造はともに飛び越し無しの left-to-right 型とした。

4.2 実験項目

特徴量として、差分パラメータ無しの z_t-13D、差分パラメータ有りの z_t-26D、の 2 種類を用いて、以下の 2 つの比較実験を行なった

実験 1 ： 平滑化部分隠れマルコフモデルによる出力確率の平滑化。

PHMM において、HMM より高次元の確率を扱っているのは出力確率計算部であるので、まずは出力確率だけ平滑化する。つまり、式 (5) において、w_t は 1 で固定し、w_o を 0 から 1 まで 0.2 刻みで変化させた。それぞれの重み毎に、モデルを学習し認識実験を行なった。

実験 2 ： 平滑化部分隠れマルコフモデルによる遷移確率の平滑化。

出力確率計算の重み (w_o) を実験 2 で最高の結果が得られる値に固定し、その時の遷移確率計算の重み (w_t) を 0 から 1 まで 0.2 毎に変化させ、実験を行なった。特徴量は z_t-13D を用いた。

4.3 実験結果

実験結果 1

出力確率を平滑化した時の認識結果を表 2、図 2 に示す。

表 2: 認識率 (%) : 出力確率の平滑化

特徴量	w_o					
	0.0	0.2	0.4	0.6	0.8	1.0
z_t-13D	98.0	98.9	99.2	99.4	99.1	98.8
z_t-26D	99.8	99.8	99.9	99.8	99.7	99.6

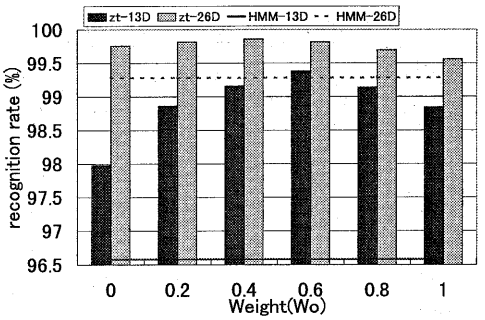

図 2: PHMM における出力確率平滑の効果

図 2 において、実線で差分無しの HMM の結果 (96.6%)、点線で差分有りの HMM の結果 (99.3%) を表している。また、重み 1 の時が PHMM の結果である。HMM と PHMM を比較すると、PHMM は差分無しの特徴量 (z_t-13D) を用いた場合で 2.2 ポイント、差分有りの特徴量 (z_t-26D) を用いた場合で 0.3 ポイント HMM の認識率を上回っている。

SPHMM は差分無しの時最高で 0.6 ポイント、差分有りの時 0.3 ポイント、PHMM を上回る結果が得られ、平滑化の効果が見られた。これは HMM と比べると差分無しで 2.8 ポイント、差分有りで 0.6 ポイント上回る結果となっている。

また、差分無しの特徴量を用いた時の認識率が、差分有りの HMM を 0.1 ポイント上回る結果となった。これは SPHMM(および PHMM) が、HMM では差分パラメタを用いないと表現できなかった過渡的な過程をも表現できることを示している。

また w_o=0 の時の出力確率は HMM と等価であり、HMM との違いは遷移確率が前出力によって条件付けられていることのみになる。そこで同特徴量を用いた HMM と比較してみると、どちらの特徴量を用いた場合も HMM を大きく上回る結果が得られていることがわかる。このことから、PHMM において遷移確率を前出力によって条件付けることの有効性がうかがえる。

実験結果 2

出力確率計算の重み (w_o) を実験 2 で最も高い認識率を得た 0.6 に固定し、遷移確率計算の重みを変化させた時の認識結果を表 3、図 3 に示す。

表 3: 認識結果：遷移確率の平滑化

特徴量	w_t					
	0.0	0.2	0.4	0.6	0.8	1.0
z_t-13D	97.7	98.7	99.0	99.1	99.0	99.4

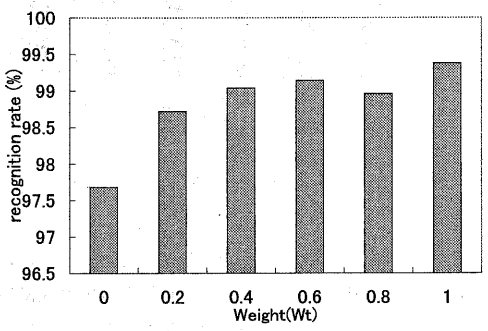

図 3: PHMM における遷移確率平滑化の効果 (z_t-13D, w_o=0.6)

重みを 1 にしたとき、つまり PHMM の遷移確率を用いた時が一番良い結果が得られ、平滑化の効果は見られなかった。このことより、過去の状態及び出力に依存して状態が遷移するという PHMM の特徴の有効性が示された。

5 まとめ

本研究では、部分隠れマルコフモデルの確率を低次の確率で平滑化した、平滑化部分隠れマルコフモデルを提案した。PHMM は通常の HMM に比べ過渡部の表現能力に優れており、出力確率だけでなく遷移確率も過去の出力に依存する点が大きな特徴であるが、HMM に比べ高次の確率を扱っているため信頼性の低下が心配された。そこで確率の平滑化を行ない、PHMM の特徴をいかしたまま、信頼性を低下させることなく計算を行なうことを可能にした。

HMM、PHMM、SPHMM を用いて実際に単語音声認識実験を行なったところ、SPHMM は遷移確率を平滑化することでは効果が見られなかったが、出力確率を平滑化することで、差分無し特徴量を用いた場合で 50.0%PHMM の誤認識率を改善することができた。これは、同特徴量を用いた場合の HMM の誤認識率を 82.4%改善する結果であり、差分有り特徴量を用いた場合の HMM の結果をも上回る結果である。また、差分有り特徴量を用いた場合 SPHMM は PHMM の誤認識率を 75.0%改善することができ、HMM の誤認識率を 85.7%改善する結果となった。

以上の結果より、平滑化部分隠れマルコフモデルの効果が示せた。

PHMM および SPHMM は、モデル内でコンテキストの影響を扱うことができる。今回の単語学習による単語音声認識ではこの特徴が活かされていないので、今後は音素学習による実験を行なっていく予定である。

参考文献

[1] 坪香 英一, 中橋 順一, "音声スペクトルの動的特徴を組み込んだ HMM," 信学論 (A), **J77-A**, 2, pp.162-172, Feb. 1994.

[2] Li Deng, M. Aksmanovic, "Speaker-Independent phonetic classification using Hidden Markov Models with mixture of trend functions" , IEEE trans on Speech and Audio Processing, vol. 5, pp.319-324, JULY. 1997.

[3] C.J.Wellekens , "Explicit correlation in Hidden Markov Model with optimal inter-frame dependence", ICASSP87, pp.383-386, 1987.

[4] 高橋敏, 松岡達雄, 南泰浩, 鹿野清宏, "フレーム間相関を利用した音韻 HMM による音声認識", 通学論 (A), **J77-A**, 2, pp.153-161, Feb. 1994.

[5] J.A.Bilmes , "Buried Markov Models for speech recognition", ICASSP99, vol.2, pp713-716, Mar. 1999.

[6] T. Kobayashi, S. Haruyama, "Partly Hidden Markov Model and its Application to Gesture Recognition", ICASSP97, vol.6, pp.3081-3084, Apr. 1997.

[7] 中川 聖一, 山本 一公, "セグメント統計量を用いた隠れマルコフモデルによる音声認識", 信学論 (D-II), **J79-D-II**, 12, pp.2032-2038, Dec. 1996.

[8] T. Kobayashi, J. Furuyama, K. Masumitsu, "Partly Hidden Markov Model and its Application to Speech Recognition",ICASSP99, vol.1, pp121-124, Mar. 1999.