時系列パターン認識のための精密な確率モデル
— 部分隠れマルコフモデルとそのジェスチャ認識・音声認識への応用—

益満 健　　古山 純子　　小林 哲則

早稲田大学 理工学部 電気電子情報工学科
〒 169　東京都新宿区大久保 3-4-1　　03-5286-3379
masuken@tk.elec.waseda.ac.jp

時系列パターンの認識手法として部分隠れマルコフモデルという新たなモデルを提案した。これは 2 重マルコフモデルの一方を隠れ状態に、他方を可観測の状態することで、left-to-right モデルの HMM が区分定常しか扱えないという問題を解決した。本モデルの性能を調べるために 2 つのシュミレーション実験を行ない、その結果 HMM に比べ過渡部の表現能力に優れていることが分かった。さらに、6 種類の手振り動作による認識実験の結果、HMM に比べエラー率を 43%改善することができた。しかし、音声での単語認識実験では HMM に及ばなかった。

時系列パターン認識, ジェスチャ認識, 音声認識,**PHMM,HMM**

More exactly random process model for time-series pattern matching
— Partly-Hidden Markov Model and Its Application to Gesture and Speech Recognition —

Ken Masumitsu　　*Junko Furuyama*　　*Tetunori Kobayashi*

Department of Electrical, Electronics & Computer Engineering
Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo, 169 Japan　　03-5286-3379
masuken@tk.elec.waseda.ac.jp

A new pattern matching method, Partly-Hidden Markov model, is proposed for time-series pattern recognition HMM, which has been used for time series pattern matting, can deal with only piecewise stationary process. We solved this problem by introducing the modified second order Markov Model, in which the first state is hidden and the second one is observable.

As the result of two simulation experimentation, the ability of expression of transient part was better than HMM. And as the result of 6 hand sign language recognition, the error rate was improved by 43% compared with HMM. However, as the result of speech recognition, the recognition rate was less than HMM. We think it is future thema.

time-series pattern matching,gesture recognition,speech recognition,PHMM,HMM

1 はじめに

統計的な時系列パターン認識のためには、確率過程のモデルが必要である。従来、この目的のためには、HMM が広く用いられてきた [1],[2]。しかしながら、より高度な認識を目指すためには、より高精度の確率過程のモデルが必要となる [3]。

これまでにも、より複雑な特徴量の変化を扱うことを目指して様々な確率モデルが提案されている。坪香らや Deng らは、セグメント内での特徴ベクトルの軌跡を多項式近似するセグメント HMM を提案している [4][5]。Wellkens や高橋らは、HMM における出力確率算出時に隣接フレームの特徴量の相関を考慮する方法を提案している [6][7]。また、Mari らは、2 重の HMM を提案し [8]、有木は、エルゴディック HMM を left-to-right に連接した 2 段階遷移構造の HMM を提案している [9]。

本研究では、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置くことで、単純な HMM より過渡部の表現能力に優れる新たな確率過程のモデルを提案する。従来の left-to-right の HMM が区分定常過程しか扱えなかったのに対し、このモデルは、left-to-right モデルであっても、次出力や状態遷移確率が 1 時刻前の出力に依存して決まるため、より複雑な振舞いを表現できる。

本報では、まず、モデルの概要を述べ、シュミレーション実験によるモデルの性能評価を行ない、最後に、ジェスチャ認識、音声認識への適用結果について述べる。

2 マルコフモデルと HMM

最も一般的な確率過程のモデルであるマルコフモデルでは、時刻 t における出力ベクトル x_tの出力確率 $P_t(x_t)$ は、過去 K 個の出力の条件付確率で与えられる。過去 K 個の出力列に、例えば S_i の状態を与えると、出力確率は次式のように表される。

$$Pr(x_t|x_{t-K}, x_{t-K+1}\cdots, x_{t-1}) = Pr(x_t|S_i) \quad (1)$$

ここで、状態 S_iは出力列に対して一意に与えられるため、確率的な揺らぎが観測系に含まれる場合、モデルの状態遷移の構造が複雑化し、単純なモデルでは観測シンボルの振舞いを表現できなくなる。

一方、HMM においては、同様な表記 $Pr(x_t|S_i)$ が用いられるが、出力列と状態の関係は一意には決めず、確率的な写像によって多数の出力列で同じ状態を共有させる。このことによって、単純なモデルで複雑な現象を扱うことが可能となる。しかしながら HMM においては、多くの出力列で共有された小数の状態だけが出力確率を決める。一つの状態に停留する限りは一定の確率分布で出力を繰り返すため、結果として区分定常の確率過程しか表現できないことになる。

3 部分隠れマルコフモデル

3.1 PHMM の概要

HMM の欠点を補うため、新たな確率過程のモデルを導入する。

まず、式 (1) の出力確率を 2 重マルコフモデルで表現する。

$$Pr(x_t|x_{t-K}\,x_{t-K+1}\cdots x_{t-1}) = Pr(x_t|S_i^f S_j^s) \quad (2)$$

ここで、 状態 S_i^f 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 は出力 x_{t-1} に対応して与える。S_i^f を F(First) 状態、S_j^s を S(Second) 状態と呼ぶことにする。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。

提案するモデルでは、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。このモデルを 部分隠れマルコフモデル (Partly-Hidden Markov Model : PHMM) と呼ぶことにする。

このモデルにおいては、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、S 状態 S_j^s すなわち x_{t-1}によって条件付けられる。このことから、モデルは区分定常以上の複雑な過程を扱うことになる。

3.2 PHMM における確率評価

PHMM を用いて出力列 $x_1x_2\cdots x_T$ が観測される確率 $Pr(x_1x_2\cdots x_T)$ を求める。

まず、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 Ps とおけばその値は、

$$\begin{aligned} P_s &= Pr(s_1^f s_1^s)Pr(x_1|s_1^f, s_1^s) \\ &\quad \times \prod_{t=1}^{T-1} Pr(s_{t+1}^f|s_t^f s_t^s)Pr(x_{t+1}|s_{t+1}^f s_{t+1}^s) \quad (3) \end{aligned}$$

定義より $s_{t+1}^f = x_t$なので、式 (3) は

$$\begin{aligned} P_s &= Pr(s_1^f, x_0)Pr(x_1|s_1^f x_0) \\ &\quad \times \prod_{t=1}^{T-1} Pr(s_{t+1}^f|s_t^f x_{t-1})Pr(x_{t+1}|s_{t+1}^f x_t) \\ &= Pr(s_1^f)Pr(x_0, x_1|s_1^f) \\ &\quad \times \prod_{t=1}^{T-1} \frac{Pr(s_{t+1}^f|s_t^f)Pr(x_{t-1}|s_{t+1}^f s_t^f)}{Pr(x_{t-1}|s_t^f)} \end{aligned}$$

$$\times \frac{Pr(x_{t+1}x_t|s^f_{t+1})}{Pr(x_t|s^f_{t+1})} \quad (4)$$

となる。求める $Pr(x_1x_2\cdots x_T)$ 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、P_s を加えればよい。

以上の議論から、PHMM のパラメータ$\boldsymbol{\theta}$は、次の 5 種類で定義できることがわかる。

・$\pi_i = Pr(s^f_1 = S^f_i)$
初期時刻に F 状態が S^f_iにある確率

・$a_{ij} = Pr(s^f_{t+1} = S^f_j | s^f_t = S^f_i)$
F 状態が S^f_iのとき、次時刻に S^f_jに遷移する確率

・$b_i(x_{t-1}) = Pr(x_{t-1}|s^f_t = S^f_i)$
F 状態が S^f_iで、S 状態 (すなわち、前時刻の出力) が x_{t-1}である確率

・$c_{ij}(x_{t-1}) = Pr(x_{t-1}|s^f_{t+1} = S^f_j, s^f_t = S^f_i)$
F 状態が S^f_iから S^f_jに遷移するときに、現在の S 状態 (すなわち、前時刻の出力) が x_{t-1}である確率

・$d_j(x_{t-1}, x_t) = Pr(x_{t-1}, x_t|s^f_t = S^f_j)$
F 状態が S^f_jであるときに、S 状態 (すなわち、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

3.3 PHMM における前向き・後向きアルゴリズム

PHMM での $Pr(x_1x_2\cdots x_T)$ の高速な求め方を考える。

時刻 t に F 状態が S^f_iにあって、x_0から x_tまでを出力する確率を前向き確率と呼ぶと、前向き確率$\alpha(i,t)$ は、次の漸化式によって与えられる。

$$\begin{aligned}\alpha(i,t) &= Pr(x_0, x_1, \cdots, x_t, s^f_t = S^f_i|\boldsymbol{\theta}) \\ &= \begin{cases} \pi_i d_i(x_0, x_1) & (t=1) \\ \sum_{j=1}^N \alpha(j,t-1) & \\ \quad \cdot \frac{a_{ji}c_{ji}(x_{t-2})}{b_j(x_{t-2})} \frac{d_i(x_{t-1},x_t)}{b_i(x_{t-1})} & (t \geq 2) \end{cases}\end{aligned}$$

ただし、N は F 状態の数である。求める $Pr(x_1x_2\cdots x_T)$ は、αを $t=1$ から順に T まで求めた後、

$$Pr(x_1x_2\cdots x_T, s^f_t = S^f_J) = \alpha(J,T)$$

により与えられる。ただし、ここでは最終状態をS^f_Jとする。

同様にして、時刻 t に F 状態が S^f_iにあって、x_{t+1}から x_Tを出力する確率として定義される、後向き確率$\beta(i,t)$ は次のように求められる。

$$\begin{aligned}\beta(i,t) &= Pr(x_{t+1}, \cdots, x_T|x_{t-1}, x_t, s^f_t = S^f_i, \boldsymbol{\theta}) \\ &= \begin{cases} 1 & (t=T) \\ \sum_{j=1}^N \frac{a_{ij}c_{ij}(x_{t-1})}{b_i(x_{t-1})} \frac{d_j(x_t,x_{t+1})}{b_j(x_t)} & \\ \quad \cdot \beta(j,t+1) & (t \leq T-1) \end{cases}\end{aligned}$$

3.4 PHMM の学習

PHMM では HMM と同様に、出力観測系列 $\boldsymbol{x}$ に対して、状態系列s^f が直接観測できない。従って、観測系列の出力確率を最大化する最適なモデルパラメータを解析的に求めることはできない。しかしながら、HMM のパラメータ推定と同様に EM アルゴリズムを適用することにより、逐次的な学習アルゴリズムを導出できる。

具体的な再推定式は次のようになる。ただし、PHMM のパラメータのうち $b_i(x_t), c_{ij}(x_t), d_i(x_{t-1}, x_t)$ は単一正規分布とした。

$$\bar{\pi}_i = \frac{Pr(\boldsymbol{x}, s^f_1 = S^f_i|\boldsymbol{\theta})}{Pr(\boldsymbol{x}|\boldsymbol{\theta})} \quad (5)$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} Pr(\boldsymbol{x}, s^f_t = S^f_i, s^f_{t+1} = S^f_j|\boldsymbol{\theta})}{\sum_{t=1}^{T-1} Pr(\boldsymbol{x}, s^f_t = S^f_i|\boldsymbol{\theta})} \quad (6)$$

$$\bar{\mu}_{b_i} = \sum_{t=1}^{T} \gamma(t,i) x_{t-1} \quad (7)$$

$$\bar{\Sigma}_{b_i} = \sum_{t=1}^{T} \gamma(t,i) \boldsymbol{C}(x_{t-1}, \mu_{b_i}) \quad (8)$$

$$\bar{\mu}_{c_{ij}} = \sum_{t=1}^{T-1} \zeta(t,i,j) x_{t-1} \quad (9)$$

$$\bar{\Sigma}_{c_{ij}} = \sum_{t=1}^{T-1} \zeta(t,i,j) \boldsymbol{C}(x_{t-1}, \mu_{c_{ij}}) \quad (10)$$

$$\bar{\mu}_{d_i} = \sum_{t=1}^{T} \gamma(t,i) y_{t-1} \quad (11)$$

$$\bar{\Sigma}_{d_i} = \sum_{t=1}^{T} \gamma(t,i) \boldsymbol{C}(y_{t-1}, \mu_{d_i}) \quad (12)$$

ただし、

$$\gamma(t,i) = \frac{Pr(\boldsymbol{x}, s^f_t = S^f_i|\boldsymbol{\theta})}{\sum_{t=1}^{T} Pr(\boldsymbol{x}, s^f_t = S^f_i|\boldsymbol{\theta})}$$

$$\zeta(t,i,j) = \frac{Pr(\boldsymbol{x}, s^f_t = S^f_i, s^f_{t+1} = S^f_j|\boldsymbol{\theta})}{\sum_{t=1}^{T-1} Pr(\boldsymbol{x}, s^f_t = S^f_i, s^f_{t+1} = S^f_j|\boldsymbol{\theta})}$$

$$\boldsymbol{C}(x,\mu) = (x-\mu)(x-\mu)^T$$

$$y_{t-1} = (x^T_{t-1}, x^T_t)^T$$

とする。

これら再推定式 (5)-(12) は、前向き確率$\alpha(i,t)$、後向き確率$\beta(i,t)$ から導かれる次の関係式を用いることにより、容易に計算することができる。

$$Pr(\boldsymbol{x}, s^f_t = S^f_i|\boldsymbol{\theta}) = \alpha(i,t)\beta(i,t) \quad (13)$$

$$Pr(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{i=1}^{N} \alpha(i,t)\beta(i,t) = \sum_{i=1}^{N} \alpha(i,T) \quad (14)$$

$$Pr(\boldsymbol{x}, s^f_t = S^f_i, s^f_{t+1} = S^f_j|\boldsymbol{\theta})$$

$$= \alpha(i,t)\beta(j,t+1)\cdot\frac{a_{ij}c_{ij}(x_{t-1})}{b_i(x_{t-1})}\frac{d_j(x_t,x_{t+1})}{b_j(x_t)} \quad (15)$$

4 シュミレーション実験

PHMM の表現能力を、シミュレーション実験によって評価する。

4.1 実験 1

本実験においては、目標値は同じであるがわたり異なる 2 つの信号を用いて、PHMM と HMM が両者をどの程度区別できるかを評価する。

区分定常的に遷移する信号として方形波を考え、滑らかに変化する過渡部を持つ信号として余弦波を考える。それぞれの学習モデルから、方形波、余弦波それぞれが生起する尤度を PHMM と HMM を用いて比較する。

4.1.1 実験条件

認識カテゴリ：

方形波　(区分定常信号)

$$y_t = \begin{cases} 20+\epsilon_t & (0 \le t \le \pi) \\ -20+\epsilon_t & (\pi < t \le 2\pi) \end{cases}$$

余弦波　(滑らかな過渡部を持つ信号)

$$y_t = 20\sin(t/2) \quad (0 \le t \le 2\pi)$$

ただし、ϵ_tは外乱であって、平均 0、分散 1 の正規分布で与えた。

データ：t に関して 0.15 毎にサンプリングし、学習用、尤度計算対象用にそれぞれ 100 データ作成。

モデルの構造：PHMM の F 状態と HMM の状態は、先端と終端に自己ループの無い飛び越しの無しの left-to-right モデル。状態数は 4 ~ 6。確率分布関数は正規分布。

特徴量：PHMM に関しては (y_t)、HMM に関しては (y_t) と $(y_t, \Delta y_t)$ を使用。ただし、Δは前時刻との差分を表す。

4.1.2 実験結果

図 1 に方形波を用いて学習を行なったモデルを用いて計算した、方形波、余弦波の平均対数尤度を示し、図 2 に余弦波を用いて学習を行なったモデルでの、平均対数尤度を示す。

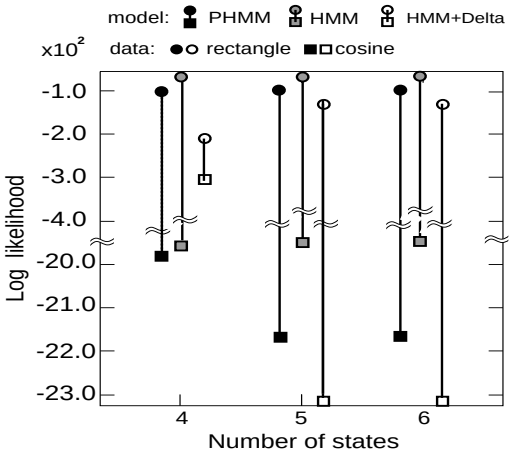

図 1: 方形波データで学習時の両モデルの平均尤度の比較

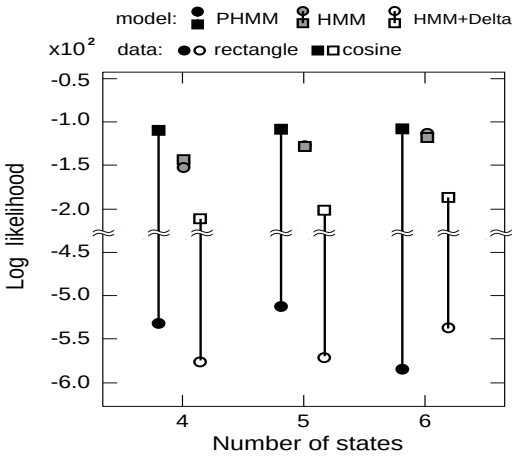

図 2: 余弦波データで学習時の両モデルの平均尤度の比較

ここでは、学習と尤度計算の対象が一致するとき、尤度が高く、一致しない時、尤度が低く抑えられることが望まれる。

方形波で学習したモデルの結果である図 1 からは、PHMM、HMM、HMM+Δともに、方形波の尤度と余弦波の尤度の差は大きく、良好な性質を保っている。一方、余弦波で学習したモデルの結果である図 2 からは、PHMM と HMM+Δでは、余弦波と方形波の尤度差が明確なのに対し、HMM

では、学習カテゴリと異なる方形波の尤度が学習カテゴリである余弦波のそれと同等に高くなっていることがわかる。特に、状態数が高く自由度が高い時、学習データでない余弦波の尤度が学習データの方形波の尤度を越えてしまうという現象が生じている。つまり、学習データが滑らかに変化する過渡部を持つ時、PHMM と HMM+Δでは、同等な渡りを持つ信号を区分定常信号から区別できるのに対し、HMM では、この区別ができないことがわかる。

これらの HMM の欠点は 特徴量に差分を含めたモデル"HMM+Δ" より改善される。しかし、PHMM では、学習データと尤度計算対象のカテゴリが一致するとき、学習データのカテゴリによらずほぼ一定の尤度値が得られているのに対し、HMM や HMM+Δでは、学習カテゴリが区分定常 (方形波) であれば非常に良好な尤度値を示すものの、区分定常でないときは劣化する。

4.2 実験 2

本実験では、異なる 2 つの目標値に様々な初期値から漸近的に収束するデータを用いて、PHMM と HMM が目標値の違いを区別にどの程度有効かを評価する。収束した部分は定常的な振舞いを示すが、この部分が長いときと短いときの両方にについて評価を行なった。

4.2.1 実験条件

認識カテゴリ : ここで使用する 関数としては臨界 2 次曲線を用いた。臨界 2 次曲線は次式で表される関数である。

$$y_t = 2ay_{t-1} - a^2 y_{t-2} + (1 - 2a + a^2)u_{t-1} + \epsilon_t$$

ここでは $a = 0.8$ とした。また、ϵ_tは外乱であって、平均 0、分散 1 の正規分布で与えた。この関数を用いて、次の二つのカテゴリを作成する。

カテゴリ 1

$$u_t = \begin{cases} 0 & t < 1 \\ 100 & t \geq 1 \end{cases}$$

カテゴリ 2

$$u_t = \begin{cases} 0 & t < 1 \\ 50 & t \geq 1 \end{cases}$$

また、初期値 y_0は $0, 20, 40, 60, 80, 100$ とする。各カテゴリの初期値毎の波形例を図 3、4 に示す。

データ : 過渡部と定常部が $0 \leq t \leq 30$ をサンプリングした過渡部と定常部から構成される 30 フレームモデルと、$0 \leq t \leq 10$ までサンプリングした過渡部のみで構成される 10 フレームモデルを学習用、尤度計算対象用にそれぞれ各初期値にごとに 50 データ、計 300 データ作成。

モデルの構造 : PHMM の F 状態と HMM の状態は、先端と終端に自己ループの無い飛び越しの無しの left-to-right モデル。状態数は 3 ~ 5。確率分布関数は正規分布。

特徴量 : PHMM に関しては (y_t)、HMM に関しては (y_t) と $(y_t, \Delta y_t)$ を使用。ただし、Δは前時刻との差分を表す。

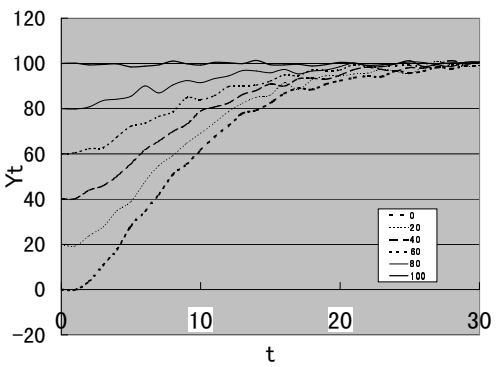

図 3: カテゴリ 1 の波形例 (初期状態は 0,20,40,60,80,100 の 6 種類)

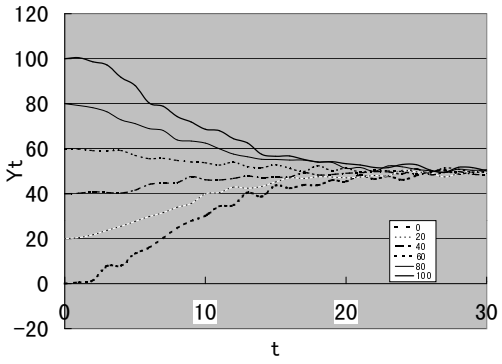

図 4: カテゴリ 2 の波形例 (初期状態は 0,20,40,60,80,100 の 6 種類)

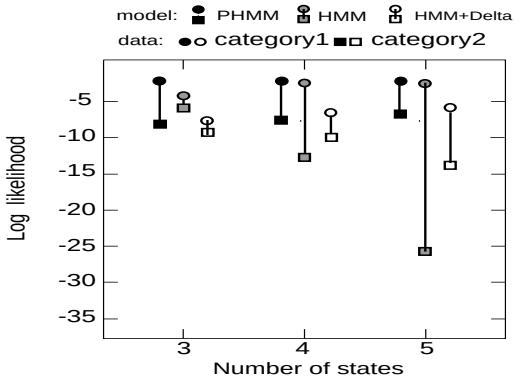

図 5: カテゴリ 1 で学習時の両モデルの平均尤度の比較 (30 フレームモデル)

4.2.2 実験結果

30 フレームモデル時の実験結果を図 5、図 6 に示し、10 フレームモデル時の実験結果を図 7、図 8 に示す。

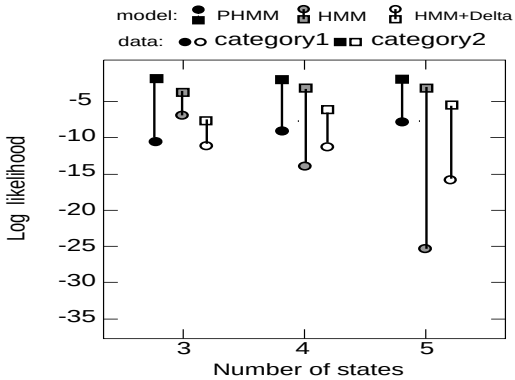

図 6: カテゴリ 2 で学習時の両モデルの平均尤度の比較 (30 フレームモデル)

全体を通じて言えることは、PHMM の与える正解カテゴリに対する尤度は常に高い値で一定に安定している。

30 フレームモデル結果 (図 5、図 6) からは、両カテゴリの尤度差は状態数が少ない時は PHMM が大きいが、状態数が増えるにしたがって、HMM、HMM+Δの尤度差が大きくなってきている。一方、10 フレームモデルの結果 (図 7、図 8) からは、PHMM は、尤度差が明確なのに対し、HMM や HMM+Δではその差は非常に小さい。

この結果から、30 フレームモデルは定常部が多

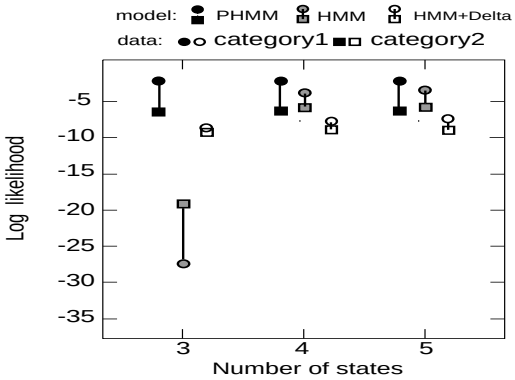

図 7: カテゴリ 1 で学習時の両モデルの平均尤度の比較 (10 フレームモデル)

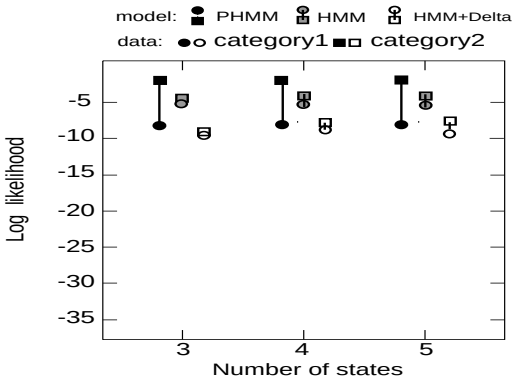

図 8: カテゴリ 2 で学習時の両モデルの平均尤度の比較 (10 フレームモデル)

いが、10 フレームモデルはほとんど定常部が無い。従って、PHMM は過渡部が多い時にその性能を発揮するが、定常部が多い時は HMM の方が優れていることが分かる。

5 PHMM のジェスチャ認識に対する応用

本手法の有効性を示すために、時系列パターン認識の一つであるジェスチャ認識に適用する。

5.1 実験条件

認識カテゴリ： 右手を使い表現する紛らわしい 6 単語。図 9 にこれら 6 種類のジェスチャを示す。

データ： 40 人の被験者が 6 語を各 4 回演じたものをデータとして収集。撮影範囲は上半身から頭部までとした。取得画像は、1 フレーム

図 9: 認識対象となる 6 種類のジェスチャ

あたり、解像度 160 × 120、RGB 各 8 ビット階調とした。これを、JPEG 圧縮 60%で、毎秒 15 フレーム取り込んだ。

本研究の目的は、時系列処理の優劣を調べることにあるため、特徴抽出段階での誤りは、抑えられることが望まれる。このため、実験では、被験者に色マーカとして赤色の手袋を装着させた。また、特徴抽出をさらに容易にするため、被験者には長袖の黒の服を着用させ、背景は無地暗色とした。

モデルの構造： PHMM の F 状態と HMM の状態は、先端と終端に自己ループの無い飛び越しの無しの left-to-right モデル。確率分布関数は正規分布。

特徴量： 原画像から、色情報と動き情報を用い、顔重心座標、右手重心座標、右手面積を抽出し、右手と顔との相対ベクトル (x_t, y_t)、速度ベクトル $(\Delta x_t, \Delta y_t)$, 右手面積の変化量 (Δs_t) 計算し、

$$z_t\text{-5D} = (x_t, y_t, \Delta x_t, \Delta y_t, \Delta s_t)$$
$$z_t\text{-3D} = (x_t, y_t, \Delta s_t)$$

を特徴量とする。

テストセットと学習セット： 収集した 4 セットのうち 1 セットをテスト用、残り 3 セットを学習用とし、組み合わせを変えて、4 回行う。

5.2 実験結果

実験結果を図 10、11 に示す。

最良の認識結果は PHMM-3D-13S(特徴量 z_t-3D、状態数 13) での 98.3%で、これは HMM での最良の結果である HMM-5D-9S、HMM-5D-11S よりも 1.5 ポイント上回っている。これは 47%の誤認識率改善に当たり、この結果は有意水準 99%で有意である。

差分の無い z_t-3D を特徴量として用いた場合、PHMM での認識率は、HMM に比べ、7 から 9 ポイント良い。HMM が区分定常過程しか扱えないという問題が現れた結果と考えられる。

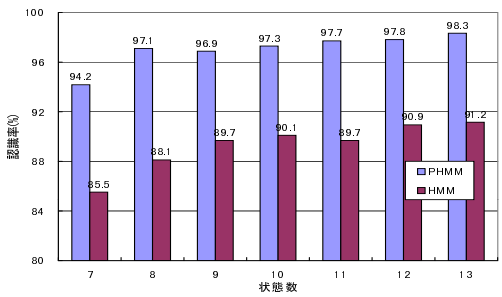

図 10: 特徴量として z_t-3D を用いたときの 6 手話認識結果

6 PHMM の音声認識に対する応用

次に、音声認識に適用する。

6.1 実験条件

データ： 学習に JNAS 男性 100 人分、認識に ATR の 216 単語セット 5 人分。学習は、HMM により自動ラベリングを行った結果を用いて、ラベル付けデータで学習を行なった。

モデルの構造： 1 音素当たり 3 状態。音素モデルを連結して単語モデルを合成。

特徴量：

$$z_t\text{-13D} = (\text{MFCC}, \text{パワー})$$
$$z_t\text{-26D} = (\text{MFCC}, \text{パワー}, \Delta\text{MFCC}, \Delta\text{パワー})$$

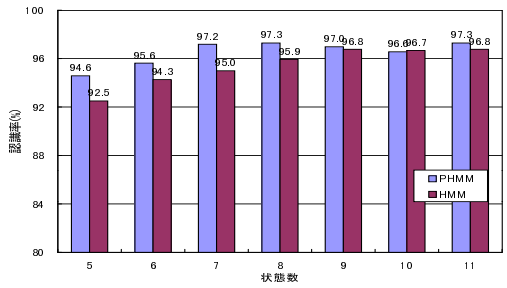

図 11: 特徴量として z_t-5D を用いたときの手話認識結果

の 2 種類の特徴量を用意する。

6.2 実験結果

音声認識実験結果を表 1 に示す。

表 1: 音声認識結果

	認識率 (%)
PHMM-13D	88.70
PHMM-26D	78.43
HMM-13D	91.02
HMM-26D	95.65

表 1 を見ると PHMM の認識率が HMM のものよりも悪くなってしまっている。この原因としては音声は過渡部のみでなく、定常部が多く含まれているために、シュミレーション実験 2 みたように、定常部の多いデータに対しては HMM の方が有効であるという結果や、PHMM はより精密なモデルを表しているので単一分布では表現しきれないということ、HMM によるラベリング結果を用いて学習したこと、等が影響していると考えられる。

7 まとめ

時系列パターン認識のための新たな確率モデルとして、部分隠れマルコフモデル (PHMM) を提案した。PHMM の特徴を調べるために 2 種類のシュミレーション実験をし、ジェスチャ認識と音声認識にこれを適用し、その有効性を確かめた。

シュミレーション実験の結果、PHMM は正解カテゴリデータについて安定して高い尤度を与えることがわかった。このことは、スポッティング認識等に PHMM が有効となる可能性を示唆する。不正解カテゴリとの尤度差については、PHMM が認識対象に依存せず、安定して良好な結果を与えるのに対し、HMM では認識対象による差が激しいことがわかった。特に、過渡部が多く含まれる時 PHMM が有効であるが、逆に定常部が多い時 HMM が有効であることがわかった。

ジェスチャ認識実験の結果、HMM の最良の結果に比べ誤認識率を 47%改善することができ、ジェスチャ認識に対する有効性が確認できた。

また、音声認識実験の結果は HMM の結果には及ばなかった。今後、検討の余地がある。

参考文献

[1] X.D.Huang, Y.Ariki, M.A.Jack, "Hidden markov models for speech recognition", Edinburgh University Press, 1990.

[2] Starner T. and Pentland A., "Visual recognition of American Sigh languate using Hidden Markov Models," Proc. Int'l Workshop on Automatic Face and Gesture Recognition 1995, June 1995.

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

[4] 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.

[5] L. Deng, M. Aksnabovic, D. Sun and J. Wu, "Speech recognition using hidden Markov models with polinomial regressin functions as nonstationary states," IEEE Trans. on Speech and Audio Proccess, vol.2 no.4, pp.507-520, Sep. 1994.

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

[7] 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", 信学論 (A), **J77-A**, 2, pp.153-161, Feb. 1994.

[8] J-F. Mari, J-P. Haton, "Automatic word recognition based on second-oder hidden Moarkov models," IEEE Trans. on Speech and Audio Process, vol.5, n.1, Jan. 1997.

[9] Y. Ariki, "Mixture density HMMs with two-level transition", Journal of Acoustic Society Japan(E), vol.14, no.4, pp.279-280, Sep. 1993.