

部分隠れマルコフモデルとそのジェスチャの認識への応用

益 満 健[†], 小林 哲 則[†]

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を提案し、ジェスチャ認識に適用した。提案するモデルは、2つの状態系列を有するマルコフモデルを基礎として、その一方を隠れ状態に、他方を可観測な状態とすることで、left-to-right型のHMMが区分定常なモデルしか扱えないという問題を解決したものである。通常セグメントモデルや 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、本モデルは、過去の出力列に関する条件付きの出力確率を扱う点、および出力確率のみならず、状態遷移も過去の出力に依存する点が大きく異なる。2種類のシミュレーション実験、および24種類の手振り動作による認識実験の結果、本手法の有効性が確認できた。

Partly-Hidden Markov Model and Its Application to Gesture Recognition

KEN MASUMITSU[†] and TETSUNORI KOBAYASHI[†]

A new pattern matching method, Partly-Hidden Markov model, is proposed and it is applied for gesture recognition. HMM, which is widely used for time series pattern matting, can deal with only piecewise stationary process. We solved this problem by introducing a modified second order Markov Model, in which the first state is hidden and the second one is observable. Conventional segment models and HMM with Δ parameters deal with joint probability of output sequence, while PHMM deal with conditional probability for given context. Furthermore, state transition probability also depends on the context. Pre-experimental results show that PHMM can deal with transient process. The result of 24 hand sign language recognition shows that the error rate was improved by 2.1 point compared with HMM.

1. はじめに

音声認識や、ジェスチャ認識などのアプリケーションで用いられる時系列パターン認識手法としては、DP マッチング^{1),2)} や隠れマルコフモデル (Hidden Markov Model: HMM) が一般的に用いられている^{3),4)}。DP マッチングはテンプレートマッチング法の一つで、時系列パターンを非線形伸縮しながら標準パターンとの比較・照合をする方法であり、時間軸方向の変動を吸収することができる。一方、HMMは確率モデルを用いたマッチング法で、音声認識、ジェスチャ認識などのアプリケーションにおいて、特に不特定者を対象にした場合、DP マッチングに比べ優位であることが確認されている⁵⁾。

しかしながら、通常用いられる left-to-right 型の

HMMは、次のような問題点が指摘されている。

- 状態数はたかだか有限個であり、同時に各状態ごとに出力確率は一定なので、全体として区分定常な確率過程しか扱うことができない。
- 状態遷移確率は、状態のみによって決まるため、それぞれの状態に停留しやすいかどうかを先験的に与えているにすぎず、出力に依存した状態遷移の振舞いを実現できない。

以上のような問題点を改善するために、本研究では、HMMに比べより動的な特徴量変化の扱いに優れた精度の高い確率モデルを提案し、その性質を2つのシミュレーション実験と、時系列パターン認識アプリケーションの1つであるジェスチャ認識実験とによって明らかにすることを試みる。

これまでに、より複雑な特徴量の時間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^{6),7)}。Wellekens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

[†] 早稲田大学理工学部

School of Science and Engineering, Waseda University
現在、日本 IBM 株式会社東京基礎研究所
Presently with IBM Research, Tokyo Research Laboratory, IBM Japan, Ltd.

る方法を提案している^{8),9)}。また, Mariらは, 2重のHMMを提案し¹⁰⁾, Arikiは, エルゴディックHMMをleft-to-rightに接続した2段階遷移構造のHMMを提案している¹¹⁾。

本研究では, 2つの状態系列を用いて現象を表現する確率モデルを基礎として, 一方を隠れ状態に, もう一方を可観測な状態におくことにより, 確率過程の表現を高精度化することを試みる。通常のセグメントモデル¹²⁾や, 特徴量の差分パラメータを利用する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, 本モデルは, 過去の出力列に関する条件付きの出力確率を扱う点, および, 出力確率のみならず状態遷移確率も過去の出力に依存して決まる点が大きく異なる。また, MariやArikiらのモデルが2重構造の両方をHMMでおいたため, 結局は大きなHMMと等価な構造となり, 前出力に対する次出力の依存関係を少数の離散状態ごとにしか表現できないのに対し, 我々のモデルでは, 一方をマルコフモデルで置くことによって, この依存関係を連続状態ごとに表現できる点も特徴である。

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 の出力確率は, 過去 K 個の出力の条件付き確率で与えられる。過去 K 個の出力列に, たとえば 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。

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

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, 単純なモデルでは多様な出力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少数の状態だけが出力確率を決める。1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。また, 状態遷移確率は状態のみによって決められるため, それぞれの状態に停留しやすいかどうかについての先験的な情報を与えているにすぎない。

HMMにおける, 状態および, 出力列の関係を図1

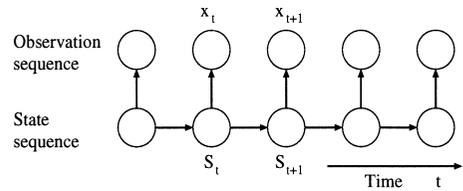


図1 HMMにおける状態・出力依存関係図
Fig. 1 Dependency graph of HMM.

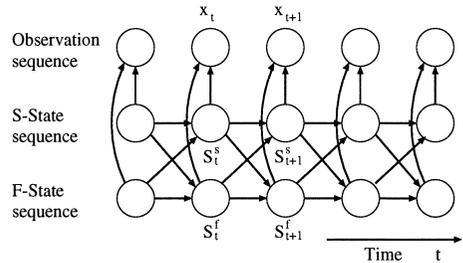


図2 2重HMMにおける状態・出力依存関係図
Fig. 2 Dependency graph of 2nd order HMM.

表1 2重マルコフモデルの分類
Table 1 Classification of 2nd order Markov Model.

	1st state(s_t^f)	2nd state(s_t^s)	
T1	observable	observable	Markov Model
T2	hidden	hidden	HMM
T3	hidden	observable	PHMM

に示す。

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, 遷移確率・出力確率が2つの状態によって条件付けられる次のような2重構造の確率モデルを考える。

出力確率: $Pr(x_t | s_t^f, s_t^s)$

遷移確率: $Pr(s_{t+1}^f | s_t^f, s_t^s), Pr(s_{t+1}^s | s_t^f, s_t^s)$

2つの状態 s_t^f, s_t^s をそれぞれ F (First) 状態, S (Second) 状態と呼ぶことにする。このモデルの状態・出力依存関係図を図2に示す。

これらの2つの状態が可観測か隠れているかの組合せにより, この2重モデルは表1のように分類することができる。2つの状態がともに可観測であるならば (T1) マルコフモデルと等価となり, 逆にともに隠れているならば (T2) HMM と等価になる。

提案するモデルは2つの状態のうち, F状態を隠れ状態にS状態を可観測にしたものである (T3)。こ

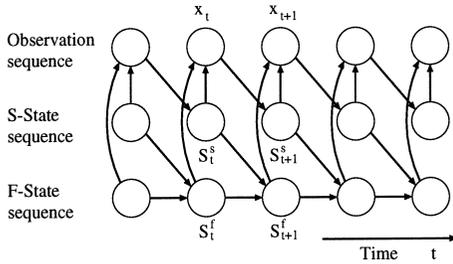


図3 PHMMにおける状態・出力依存関係図 ($s_t^s = x_{t-1}$)
Fig. 3 Dependency graph of PHMM ($s_t^s = x_{t-1}$).

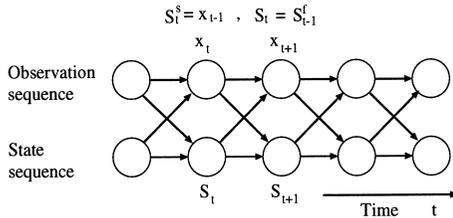


図4 PHMMにおける状態・出力依存関係図 ($s_t^s = x_{t-1}$)
Fig. 4 Dependency graph of PHMM ($s_t^s = x_{t-1}$).

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。S 状態は、過去のどの部分の出力列から観測可能とするかについては任意性があるが、本報では最も単純なモデルとして、前出力のみから観測できるものを扱う。したがって、図2の状態・出力依存関係図は、図3のように書き直され、さらに $s_t^s = x_{t-1}$ として S 状態の表記を除けば、PHMM の状態・出力依存関係図は図4のようになる。以下、図4の形の PHMM について検討を進める。

PHMM においては、出力確率・状態遷移確率を決める条件部の半分 (F 状態) が、多くの出力列で共有されることから、単純なマルコフモデルに比べればモデルの複雑さを抑えることができる。また、それらの確率は、少数の隠れ状態のみでなく、連続空間に定義された S 状態、すなわち前時刻出力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表現できることになる。

3.2 PHMM における確率評価

PHMM を用いて出力列 $x_1 \cdots x_T$ が観測される確率 $Pr(x_1 \cdots x_T)$ を求める。

2重の HMM において出力列 $x_1 \cdots x_T$ が観測される確率 $Pr(x_1 \cdots x_T)$ は、

$$\begin{aligned} Pr(x_1 \cdots x_T) &= \sum_{\text{all } s^f, s^s} Pr(s_1^f s_1^s) Pr(x_1 | s_1^f s_1^s) \\ &\times \prod_{t=1}^{T-1} Pr(s_{t+1}^f | s_t^f s_t^s) Pr(s_{t+1}^s | s_t^f s_t^s) \\ &\times Pr(x_{t+1} | s_{t+1}^f s_{t+1}^s) \end{aligned} \quad (2)$$

と表せる。

一方、PHMM においては、S 状態が可観測であるために、

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

となる。さらに、

$$s_{t+1}^s = x_t \quad (4)$$

なる関係式を用いれば、 $Pr(x_1 \cdots x_T)$ は以下のように表せる。

$$\begin{aligned} Pr(x_1 \cdots x_T) &= \sum_{\text{all } s^f} Pr(s_1^f x_0) Pr(x_1 | s_1^f x_0) \\ &\times \prod_{t=1}^{T-1} Pr(s_{t+1}^f | s_t^f x_{t-1}) Pr(x_{t+1} | s_{t+1}^f x_t) \end{aligned} \quad (5)$$

通常のセグメントモデルや Δ パラメータを含む確率過程のモデルが、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、提案モデルは、過去の出力列に関する条件付きの出力確率を扱う点、および、出力確率のみならず状態遷移も過去の出力に依存する点が大きく異なる。

ベイズの定理より式 (5) は、

$$\begin{aligned} Pr(x_1 \cdots x_T) &= \sum_{\text{all } s^f} Pr(s_1^f) Pr(x_0, x_1 | s_1^f) \\ &\times \prod_{t=1}^{T-1} \frac{Pr(s_{t+1}^f | s_t^f) Pr(x_{t-1} | s_t^f s_{t+1}^f)}{Pr(x_{t-1} | s_t^f)} \\ &\times \frac{Pr(x_t x_{t+1} | s_{t+1}^f)}{Pr(x_t | s_{t+1}^f)} \end{aligned} \quad (6)$$

となることから、PHMM は次の 5 種類のパラメータで表現できることが分かる。

- $\pi_i = Pr(s_1^f = S_i^f)$
初期時刻に F 状態が S_i^f にある確率
- $a_{ij} = Pr(s_{t+1}^f = S_j^f | s_t^f = S_i^f)$
F 状態が S_i^f のとき、次時刻に S_j^f に遷移する確率
- $b_i(x_{t-1}) = Pr(x_{t-1} | s_t^f = S_i^f)$
F 状態が S_i^f で、S 状態 (すなわち、前時刻の出

力)が x_{t-1} である確率

- $c_{ij}(x_{t-1}) = Pr(x_{t-1}|s_t^f = S_j^f s_{t+1}^f = S_i^f)$
F 状態が S_i^f から S_j^f 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
- $d_j(x_{t-1}, x_t) = Pr(x_{t-1}x_t|s_t^f = S_j^f)$
F 状態が S_j^f 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

$Pr(x_1x_2 \cdots x_T)$ の高速な求め方を考える. 時刻 t に F 状態が S_i^f 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, PHMM のモデルパラメータを θ とするとき, 次の漸化式によって与えられる.

$$\begin{aligned} \alpha(i, t) &= Pr(x_0x_1 \cdots x_t s_t^f = S_i^f | \theta) \\ &= \begin{cases} \pi_i d_i(x_0, x_1) & (t=1) \\ \sum_{j=1}^N \alpha(j, t-1) \frac{a_{ji} c_{ji}(x_{t-2})}{b_j(x_{t-2})} \frac{d_i(x_{t-1}, x_t)}{b_i(x_{t-1})} & (t \geq 2) \end{cases} \end{aligned} \quad (7)$$

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,

$$Pr(x_1x_2 \cdots x_T) = \sum_{i=1}^N \alpha(i, T) \quad (8)$$

により与えられる.

同様に, 時刻 t に F 状態が S_i^f 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.

$$\begin{aligned} \beta(i, t) &= Pr(x_{t+1} \cdots x_T | x_{t-1} x_t s_t^f = S_i^f | \theta) \\ &= \begin{cases} 1 & (t=T) \\ \sum_{j=1}^N \frac{a_{ij} c_{ij}(x_{t-1})}{b_i(x_{t-1})} \frac{d_j(x_t, x_{t+1})}{b_j(x_t)} \cdot \beta(j, t+1) & (t \leq T-1) \end{cases} \end{aligned} \quad (9)$$

3.4 PHMM と HMM のモデルの自由度について

left-to-right 型の PHMM と HMM 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, 1 つの確率分布を単一の正規分布で表現するときの 1 状態あたりパラメータの数を表 2 に示す. ただし, 表は特徴量の各次元相互の相関を考慮するとき (フルコバリアンスの共分散行列を用いるとき) の値である.

表 2 PHMM と HMM の 1 状態あたりの自由度の比較 (D : 特徴量の次元数)

Table 2 Comparison of degree of freedom per state between PHMM and HMM (D : Feature Vector Dimension).

	Degree of Freedom
PHMM	$3D^2 + 6D$
HMM	$\frac{1}{2}D^2 + \frac{3}{2}D$

表 2 より, 大雑把にいうと, PHMM は HMM に比べ, 6 倍程度自由度が大きいことになる.

PHMM はこのように高い自由度を持つため, 記述能力が増える反面, 学習に必要なデータ数が増える可能性を持っている.

3.5 パラメータ推定

PHMM では HMM と同様に, 出力列 x ($= x_1, x_2, \dots, x_T$) から, 状態系列 s^f ($= s_1^f, s_2^f, \dots, s_T^f$) を直接観測できない. したがって, 観測系列の出力確率 $Pr(x|\theta)$ を最大化する最適なモデルパラメータ θ を解析的に求めることはできない. しかし, EM アルゴリズムとして知られるような繰返し手法などによって, $Pr(x|\theta)$ が局所的に最大になるモデルパラメータ θ を求めることができる.

各パラメータを単一正規分布としたときの再推定式を以下に示す (具体的な導出過程は付録 A.1 に記す).

$$\bar{\pi}_i = \frac{Pr(x s_1^f = S_i^f | \theta)}{Pr(x | \theta)} \quad (10)$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} Pr(x s_t^f = S_i^f s_{t+1}^f = S_j^f | \theta)}{\sum_{t=1}^{T-1} Pr(x s_t^f = S_i^f | \theta)} \quad (11)$$

$$\bar{\mu}_{b_i} = \sum_{t=1}^T \gamma(t, i) x_{t-1} \quad (12)$$

$$\bar{\Sigma}_{b_i} = \sum_{t=1}^T \gamma(t, i) C(x_{t-1}, \mu_{b_i}) \quad (13)$$

$$\bar{\mu}_{c_{ij}} = \sum_{t=1}^{T-1} \zeta(t, i, j) x_{t-1} \quad (14)$$

$$\bar{\Sigma}_{c_{ij}} = \sum_{t=1}^{T-1} \zeta(t, i, j) C(x_{t-1}, \mu_{c_{ij}}) \quad (15)$$

$$\bar{\mu}_{d_i} = \sum_{t=1}^T \gamma(t, i) y_{t-1} \quad (16)$$

$$\bar{\Sigma}_{d_i} = \sum_{t=1}^T \gamma(t, i) C(y_{t-1}, \mu_{d_i}) \quad (17)$$

ただし,

$$\gamma(t, i) = \frac{Pr(\mathbf{x}s_t^f = S_i^f | \theta)}{\sum_{t=1}^T Pr(\mathbf{x}s_t^f = S_i^f | \theta)}$$

$$\zeta(t, i, j) = \frac{Pr(\mathbf{x}s_t^f = S_i^f s_{t+1}^f = S_j^f | \theta)}{\sum_{t=1}^{T-1} Pr(\mathbf{x}s_t^f = S_i^f s_{t+1}^f = S_j^f | \theta)}$$

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

$$y_{t-1} = (x_{t-1}^T, x_t^T)^T$$

とする。

ここで、再推定導出式 (10) ~ (17) 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。

$$Pr(\mathbf{x}s_t^f = S_i^f | \theta) = \alpha(i, t)\beta(i, t)$$

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

$$Pr(\mathbf{x}s_t^f = S_i^f s_{t+1}^f = S_j^f | \theta) = \alpha(i, t)\beta(j, t+1) \cdot \frac{a_{ij}c_{ij}(x_{t-1})}{b_i(x_{t-1})} \frac{d_j(x_t, x_{t+1})}{b_j(x_t)}$$

4. シミュレーション実験

PHMM の表現能力を、2 種類のシミュレーション実験により評価した。

4.1 実験 1

区分定常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た。これらに雑音を加えた時系列データを用いて PHMM, HMM を学習し、それぞれのモデルから、方形波, 余弦波が生起する尤度を比較した。これらの比較によって、それぞれの確率モデルが、始末端における値が同じで渡りの異なる信号を区別できるかどうかを調べた。

4.1.1 実験条件

認識カテゴリ

方形波 (区分定常信号)

$$y = \begin{cases} 20 + N(0, 1) & (0 \leq x \leq \pi) \\ -20 + N(0, 1) & (\pi < x \leq 2\pi) \end{cases}$$

余弦波 (滑らかな過渡部を持つ信号)

$$y = 20 \cos(x/2) + N(0, 1) \quad (0 \leq x \leq 2\pi)$$

ただし、 $N(0, 1)$ は平均 0、分散 1 の正規分布に従う乱数を示す。両カテゴリとも学習用、尤度計算対象用にそれぞれ 100 データを作成して用いた。

モデルの構造

PHMM の F 状態は飛び越しなしの left-to-right モデルとした。状態数は 2~4 とし、確率分布関数は単

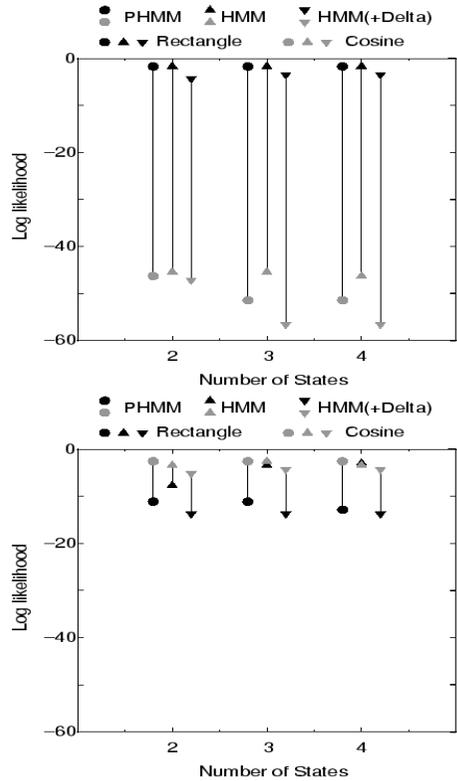


図 5 各カテゴリデータの平均スコアの比較 (上段: 方形波モデル, 下段: 余弦波モデル)

Fig. 5 Comparison of average score (upper: rectangular model, lower: cosine model).

一正規分布とした。

比較項目

各カテゴリごとにモデルパラメータを推定し、それぞれのカテゴリのデータの対数尤度 $Pr(x|\theta)$ を求めた。同条件の HMM, および、特徴量に差分値を付加した HMM と比較した。

4.1.2 実験結果

図 5 上段に方形波を用いて学習を行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 5 下段に余弦波を用いて学習を行ったモデルでの平均対数尤度を示す。

プロットの形の違いは、用いたモデルの違いを示している。また、色が黒いプロットは、尤度計算対象データが方形波データであることを示し、灰色のプロットは、それが余弦波データであることを示す。

横軸は状態数を示し、縦軸は各尤度計算対象データにおける尤度の平均を示している。学習と尤度計算の対象データが一致するとき尤度が高く、一致しないとき、尤度が低く抑えられることが望まれるため、結果として、両者の差が大きいとき (縦軸に関して離れている

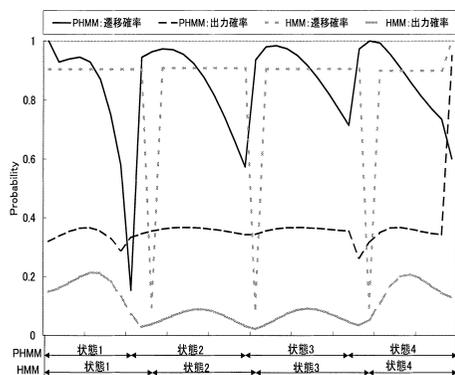


図 6 各時刻ごとの遷移・出力確率の変化(学習:余弦波, 評価:余弦波)

Fig. 6 Example of values of transition and output probabilities (cosine data).

とき)ほど, 識別能力が高いということを示している.

方形波で学習したモデルにおいては, HMM, PHMM とともに, 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ことが分かる. 状態数による違いも少ない. しかしながら, 余弦波で学習したモデルを用いたとき, PHMM では余弦波と方形波の尤度差が明確なのに対し, HMM では本来低くあるべき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分かる. つまり, 学習データが滑らかに変化する過渡部を持つとき, PHMM 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分かる.

この結果について考察するために, 余弦波学習モデルを用いて, 余弦波データを評価したときの各時刻の遷移・出力確率の変化の様子を表したものを図 6 に示す. 値は, 最適パス上の状態遷移における状態遷移確率と出力確率を表したものである. モデルの状態数は 4 である.

left-to-right の HMM では, 状態遷移確率は, 現状態にとどまるときの値と, 次状態へ進むときの値の 2 通りしかなく, 一般にこれらの比はどの状態でも大きく変わらない. このため, 現状態にとどまる限り一定の遷移確率となり, 次状態に進むときに一度だけ異なる値をとることになる. よって, どこで状態の切り替わりが生じて, 状態遷移確率の累積値には大きな差が生じない. このため, 状態遷移確率は, 状態の切り替わり時刻を決めるのに大きな影響力を持たず, 結果として, HMM では出力確率の累積値のみを最適化する方向で状態の切り替わり時刻を決める傾向がある. 図 6 から, HMM では出力確率の時間変化が滑らかなこ

とが見てとれるが, これは各時刻ごとに出力の高い方の状態を選びながら遷移が行われた結果としてみることができる. 他方, PHMM においては, 状態遷移が前状態および前出力によって決まるために状態遷移確率は動的に変化し, どこで状態が切り替わるかについて, 大きな影響力を持つことになる. 図 6 から, PHMM では状態切替え時刻において出力確率が不連続な値の変化をしていることが見てとれるが, これは状態遷移確率を高くしようとした結果, 出力確率の低い状態をあえて選択した結果としてみることができる. 要するに, HMM では出力確率だけを高くするよう, 状態を選択的に移動する可能性があるのに対し, PHMM では, 状態遷移確率が前時刻の出力に依存して決まることの影響で, そのような状態遷移を生じる可能性が HMM に比して小さい. 図 5 に示した結果において, HMM で方形波を余弦波から区別できず, PHMM でそれができたのは, HMM では余弦波のモデルの一部の状態のみを選択的に使う(出力確率の累積値を高くする観点から都合の悪い一部の状態をスキップしてしまう)ことによって, 不当に高い尤度を与えるのに対し, PHMM ではそのような遷移が許されなかった結果と考えることができる.

1 章で紹介した従来研究はすべて HMM 型の状態遷移確率を採用している. そのため, それらは状態遷移確率に関して, すべて HMM と同様な弱点を持っていることになる.

4.2 実験 2

異なる 2 つの目標値に様々な初期値から漸的に収束するデータを用いて, PHMM と HMM が目標値の違いの区別にどの程度有効かを評価した. 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についても評価を行った.

4.2.1 実験条件

認識カテゴリ: ここで使用する信号としては, 次式で表される臨界 2 次曲線を用いた.

$$y_x = 2ay_{x-1} - a^2y_{x-2} + (1-2a+a^2)u_{x-1} + N(0, 1)$$

ここでは, $a = 0.8$ とした. この関数を用いて, 次の 2 つのカテゴリを作成した.

カテゴリ 1

$$u_x = \begin{cases} 0 & x < 1 \\ 100 & x \geq 1 \end{cases}$$

カテゴリ 2

$$u_x = \begin{cases} 0 & x < 1 \\ 50 & x \geq 1 \end{cases}$$

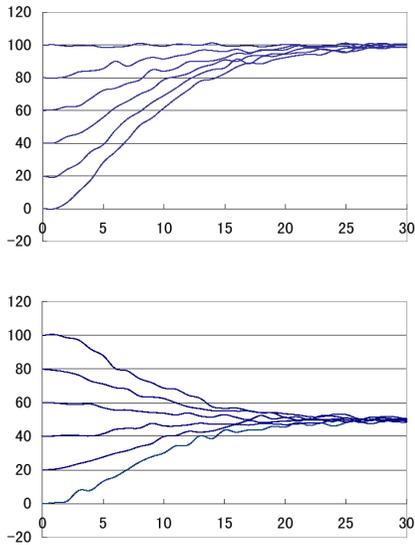


図 7 シミュレーション実験 2 に用いた波形例(上段: カテゴリ 1, 下段: カテゴリ 2)

Fig. 7 Examples of signals used for 2nd simulation experiment (upper: category1, lower: category2).

また, 初期値 y_0 は 0, 20, 40, 60, 80, 100 とする. 両カテゴリの各初期値ごとの波形例を図 7 に示す.

データ: $0 \leq t \leq 30$ のデータを用いた過渡部と定常部から構成される 30 フレームモデルと, $0 \leq t \leq 10$ のデータを用い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ごとに 50 データ, 計 300 データ作成した.

モデルの構造: PHMM の F 状態は飛び越しなしの left-to-right モデルとした. 状態数は 2~4, 確率分布関数は単一正規分布とした.

比較項目: 各カテゴリごとにモデルパラメータを推定し, それぞれのカテゴリのデータの対数尤度(スコア) $Pr(x|\theta)$ を求めた. 同条件の HMM, および, 特徴量に差分値を付加した HMM と比較した.

4.2.2 実験結果

図 8 上段に 30 フレームモデルのカテゴリ 1 を用いて学習を行ったモデルと, カテゴリ 2 を用いて学習を行ったモデル, それぞれによる正解カテゴリ・不正解カテゴリの平均対数尤度を示す. また, 図 8 下段に 10 フレームモデルの平均対数尤度を示す. なお, 本実験においては, カテゴリ 1 で学習したモデルによる結果と, カテゴリ 2 で学習した際の結果に大きな差異がみられなかったため個別に示さずに, 両者の平均を示

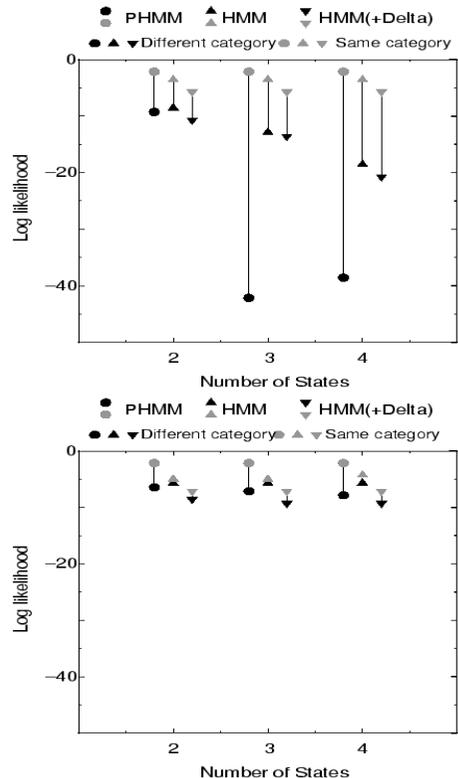


図 8 各カテゴリデータの平均スコアの比較(上段: 30 フレームモデル, 下段: 10 フレームモデル)

Fig. 8 Comparison of average score (upper: 30 frame model, lower: 10 frame model).

した.

全体を通じて, HMM に比べて, PHMM は尤度差が大きく, 目的値の識別を初期値に依存しないモデルによって行う能力が, PHMM では HMM に比べて高いことが分かる. さらに, PHMM の与える正解カテゴリに対する尤度がつねに高い値で一定に安定しているということも分かる.

5. ジェスチャ認識実験

5.1 実験に用いるデータ

本実験において識別対象としたジェスチャは, 表 3 に示す 24 種類である. 腕全体による比較的大きい動作を行うジェスチャが選ばれている. これらの 24 種類のジェスチャ 35 人分を各 4 セット用意した. 例として「左回り」のジェスチャを図 9 に示す.

撮影範囲は被験者の上半身から頭部までとし, 1 フレームあたり, 解像度 160×120 , RGB 各 8 ビット階調とし, JPEG 圧縮 70% で毎秒 15 フレームで取り込んだ. また, 被験者には顔, 手以外の肌が露出しないように長袖の服を着用してもらい, 自由背景下で撮



図 9 左回りのジェスチャ例

Fig.9 Example gesture of turn left.

表 3 認識実験に使用したジェスチャ(24種類)

Table 3 The gestures for the recognition experiment.

“ダメ”, “イエス”, “万歳”, “さよなら”, “こんにちは”, “静かに”, “こんな大きさの(大)”, “こんな大きさの(小)”, “拍手”, “good”, “もっと右”, “もっと左”, “行け”, “来い”, “招く”, “止まれ”, “右回り”, “左回り”, “右”, “左”, “上”, “下”, “赤”, “黄”

影を行った。

5.2 特徴抽出アルゴリズム

1. 肌色尤度変換 取り込んだ画像を $L^*a^*b^*$ 表色系に変換する。このうちの a^*b^* は色本来の情報を表し、照明の影響を受けにくいという特徴を持つ。この性質を利用して、 a^*b^* のもとで肌の色の分布を二次元単一正規分布であると仮定し、各画素を肌色尤度に変換し、255 階調肌色尤度 $L(i,j,t)$ 画像を作成する。この操作により、顔・手領域の肌の色領域を強調することができる。
2. 動領域抽出 肌色尤度画像には、顔・手領域以外の背景・服に存在する肌らしい色も強調されてしまう。そこで、1. と並行して過去 N フレーム間の平均画像との差分を計算することにより、動領域を抽出し、2 値動領域画像 $M(i,j,t)$ を作成する。
3. 動肌色尤度画像 肌色で、かつ動領域は手と考えられるので、画像 $M(i,j,t)$ で $L(i,j,t)$ をマスク処理をすることにより動肌色尤度画像を得る。
4. 次元圧縮 動肌色尤度画像の次元 160×120 を圧縮するために、 3×3 のメッシュ処理と DCT を行う。得られた DCT 成分の低周波成分 $z_t = 14$ 次元(ただし、直流成分を除く)を特徴量として用いる。

5.3 実験条件

学習データと認識データ 各被験者のデータを各カテゴリとも 1 つずつ含む 4 組みのデータセットを作った。そのうち、3 セットを用いて確率モデルを学習し、残りの 1 セットの認識を行った。学習セットと認識セットの組合せを変えて、4 回認識実験を行った。したがって、各カテゴリの認識データの総数は 140 であり、1 確率モデルあたりの学習データ数は 105 である。

モデルの構造 F 状態は、飛び越しなしの left-to-right モデルとし、確率分布関数は単一正規分布とした。
比較項目 F 状態の状態数を 5 から 15 まで変化させ、認識率を調べた。これを、同条件の HMM および、特徴量に差分値を付加した HMM と比較した。

5.4 実験結果

実験結果を図 10 に示す。

前述したように、PHMM と HMM における自由度は異なるため、最適な学習データ数、状態数は異なる。そこで、本実験においては状態数を変化させ、各手法ごとに最良の認識率を求め、この値によって各手法を比較する。

PHMM における最良の認識率は 83.3%、HMM における最良の認識率は 77.9%であった。HMM に関しては、特徴量に差分パラメータを加えることにより、差分を用いないときに比べ認識率は 3.3 ポイント改善され 81.2%となったが、それでも PHMM の認識率に比べて 2.1 ポイント及ばなかった。これは有意水準 99%で有意である。

また、HMM における最良の認識率は、差分特徴量を用いたときに状態数 11 で達成されるが、PHMM で

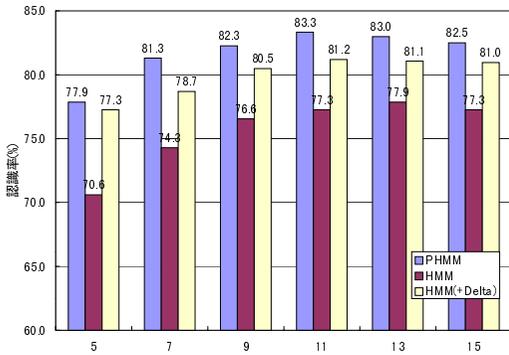


図 10 ジェスチャ認識実験結果
Fig. 10 Results of gesture recognition.

は状態数 7 で同程度の認識率を得ることができている。

以上のことから、PHMM のジェスチャ認識に対する有効性が確認された。

6. 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、これをジェスチャー認識に適用した。HMM が区分定常な確率過程しか扱えないのに対し、PHMM では出力分布確率や状態遷移確率が前出力に依存して決まるため、より複雑な確率過程を表現できる。

2 種類のシミュレーション実験を通じて、過渡部の表現能力における PHMM の HMM に対する優位性、正解カテゴリに対する尤度の安定という特徴が確認できた。

時系列パターン認識アプリケーションの 1 つである、ジェスチャ認識実験に応用した結果、HMM の最良の結果に比べ 2.1 ポイント改善することができ、PHMM が実際のアプリケーションにも有効であることが確認できた。

なお今回は、S 状態が前出力からのみに依存する構造の PHMM を扱ったが、3.1 節に述べたように、S 状態を何から観測可能とするかに応じて様々な構造の PHMM が構成できる。今後、最適な構造の決定法について検討を行いたい。

PHMM は、HMM に比べより精密な確率モデルであるから、尤度値の信頼性が向上することが期待できる。このため、尤度値による閾値設定などが必要とされるスポッティングにおいてより威力を発揮することが期待される。今後、このような観点での応用についても検討を進めたい。

参考文献

- 1) Sakoe H. and Chiba S.: Dynamic programming algorithm optimization for spoken word recognition, *IEEE Trans. Acoustics, Speech and Signal Processing*, ASSP-26, 1, pp.43-49 (Feb. 1978).
- 2) 高橋勝彦, 関進, 小島浩, 岡隆一: ジェスチャ動画のスポッティング認識, 信学論 (D-II), J77-D-II, 8, pp.1552-1561 (Aug. 1994).
- 3) 小林哲則: 隠れマルコフモデルに基づく音声認識, 電気学会論文誌, 113-C, 5, pp.259-301 (May 1993).
- 4) Starner, T. and Pentland, A.: Visual recognition of American Sign language using Hidden Markov Models, *Proc. Int'l Workshop on Automatic Face and Gesture Recognition 1995* (June 1995).
- 5) 春山智, 小林哲則: 動画処理による手振り動作認識, 信学総大'97, D-12-55 (1997).
- 6) 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).
- 7) Deng, L., Aksnabovic, M., Sun, D. and Wu, J.: Speech recognition using hidden Markov models with polynomial regression functions as non-stationary states, *IEEE Trans. Speech and Audio Process*, Vol.2, No.4, pp.507-520 (1994).
- 8) Wellekens, C.J.: Explicit correlation in Hidden Markov Model with optimal inter-frame dependence, *Proc. ICASSP87*, pp.383-386 (1987).
- 9) 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).
- 10) Mari, J-F. and Haton, J-P.: Automatic word recognition based on second-order hidden Markov models, *IEEE Trans. Speech and Audio Process*, Vol.5, No.1 (1997).
- 11) Ariki, Y.: Mixture density HMMs with two-level transition, *Journal of Acoustic Society Japan (E)*, Vol.14, No.4, pp.279-280 (1993).
- 12) 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).
- 13) Kobayashi, T. and Haruyama, S.: Partly Hidden Markov Model and its Application to Gesture Recognition, *IEEE Proc. ICASSP97*, Vol.VI, pp.3081-3084 (1997).
- 14) 伊藤政雄, 竹村茂: 手話入門, 廣済堂。

付 録

A.1 再推定式の導出過程

まず, $Pr(x|\theta)$ を最大化する θ を直接求める代わりに, 次式で表される補助関数 Q を最大化する $\bar{\theta}$ を漸次求めることを考える.

$$Q(\theta, \bar{\theta}) = \sum_{sf} Pr(x s^f | \bar{\theta}) \log Pr(x s^f | \theta) \quad (18)$$

ただし, $\bar{\theta}$ は再推定されたパラメータとする. なぜなら,

$$Q(\theta, \bar{\theta}) \geq Q(\theta, \theta) \rightarrow Pr(x|\bar{\theta}) \geq Pr(x|\theta) \quad (19)$$

という関係があるため, $\bar{\theta}$ は $Pr(x|\theta)$ の極大値を与える θ に収束する.

$\log Pr(x s^f | \theta)$ は, 3.2 節で定義したパラメータを用いるとき, 次のように表現できる.

$$\begin{aligned} \log Pr(x s^f | \theta) &= \log \pi_i + \sum_{t=1}^{T-1} \log a_{ij} \\ &+ \sum_{t=1}^{T-1} \log c_{ij}(x_{t-1}) - \sum_{t=1}^{T-1} \log b_i(x_{t-1}) \\ &+ \sum_{t=0}^{T-1} \log d_j(x_t, x_{t+1}) - \sum_{t=1}^{T-1} \log b_j(x_t) \quad (20) \end{aligned}$$

式 (20) を式 (18) に代入すると,

$$\begin{aligned} Q(\theta, \bar{\theta}) &= Q_\pi(\theta, \pi) + \sum_{i=1}^N Q_a(\theta, a_i) \\ &+ \sum_{i=1}^N \sum_{j=1}^N Q_c(\theta, c_{ij}(x_t)) \\ &+ \sum_{i=1}^N Q_{bd}(\theta, b_i(x_t), d_i(x_t, x_{t+1})) \quad (21) \end{aligned}$$

となる. ただし,

$$Q_\pi(\theta, \pi) = \sum_{i=1}^N Pr(x s_1^f = S_i^f | \theta) \log \pi_i$$

$$Q_a(\theta, a_i) = \sum_{j=1}^N \sum_{t=1}^{T-1} Pr(x s_t^f = S_i^f s_{t+1}^f = S_j^f | \theta) \times \log a_{ij}$$

$$Q_c(\theta, c_{ij}(x_t)) = \sum_{t=1}^N Pr(x s_t^f = S_i^f s_{t+1}^f = S_j^f | \theta) \times \log c_{ij}(x_t)$$

$$\begin{aligned} Q_{bd}(\theta, b_i(x_t), d_i(x_t, x_{t+1})) \\ &= \sum_{i=1}^N \sum_{t=1}^T Pr(x s_t^f = S_i^f | \theta) \\ &\times \{ \log d_i(x_t, x_{t+1}) - 2 \log b_i(x_t) \} \end{aligned}$$

である.

式 (21) の右辺は 4 つの独立な項に分かれているので, それぞれの項を個別に最大化することで $Q(\theta, \bar{\theta})$ を最大化できる. パラメータ $b_i(x_t)$, $c_{ij}(x_t)$, $d_i(x_t, x_{t+1})$ を単一正規分布とすると, PHMM の再推定モデル θ 式 (10) ~ (17) が導出される.

(平成 11 年 10 月 14 日受付)

(平成 12 年 9 月 7 日採録)



益満 健 (正会員)

1997 年早稲田大学理工学部電気電子情報工学科卒業. 1999 年同大学大学院修士課程修了. 同年日本アイ・ピー・エム (株) 入社. パターン認識・画像処理に関する研究に従事.



小林 哲則 (正会員)

1980 年早稲田大学理工学部電気電子情報工学科卒業. 1985 年同大学大学院博士課程修了. 工学博士. 同年法政大学講師. 同助教, 早稲田大学助教を経て, 現在早稲田大学理工学部電気電子情報工学科教授. 音声言語処理, 動画画像処理, 知能ロボット等の研究に従事. 電子情報通信学会, 日本ロボット学会, 人工知能学会, 日本音響学会, IEEE, ACM 等会員.