

PHMMによるジェスチャ認識

5E-10

益満 健 早田 啓介 小林 哲則
早稲田大学 理工学部

1 はじめに

今日、時系列パターン認識において、広く一般に用いられている、left-to-rightのHMMは、区分定常な確率過程しか扱うことができない。このため、より精密な認識を行なうためには、過渡部の表現に優れた確率モデルが必要とされている [1][2]。

我々は、2重マルコフモデルを基に部分隠れマルコフモデル (PHMM) を提案した。このモデルは従来提案されてきた多くのモデルを、自然な形で含んだモデルとなっている。[3]。

本稿では、PHMMの特徴をシミュレーション実験によって明らかにするとともに、ジェスチャ認識に応用し、その効果を確認する。

2 PHMMの概要

時刻 t における出力が過去 K 個の出力列 x_{t-K}, \dots, x_{t-1} 条件付き確率で表現される確率過程を、次のような二つの状態を用いた、2重マルコフモデルで表現する。

$$Pr(x_t | x_{t-K} x_{t-K+1} \dots x_{t-1}) = Pr(x_t | S_t^f S_t^s) \quad (1)$$

この二つの状態、状態 S_t^f を F 状態、 S_t^s を S 状態と呼ぶ。ここで、S 状態 (s_t^s) は、 x_{t-1} の出力に対して一意に $S_{t+1}^s = x_t$ と決める。一方、F 状態 (s_t^f) は $x_{t-K} x_{t-K+1} \dots x_{t-1}$ の出力列に対して、確率的に写像させるとし、 s_{t-1}^f と s_{t-1}^s の条件付き出力として、与えられるものとする。

このとき、出力列 $x_1 x_2 \dots x_T$ がある F 状態列と S 状態列から生起する確率は、

$$\begin{aligned} P_s &= Pr(x_1 x_2 \dots x_T s_1^f s_2^f \dots s_T^f s_1^s s_2^s \dots s_T^s) \\ &= Pr(x_1^f | s_1^f) Pr(x_1 | s_1^f s_1^s) \\ &\quad \times \prod_{t=1}^{T-1} Pr(s_{t+1}^f | s_t^f s_t^s) Pr(x_{t+1} | s_{t+1}^f s_{t+1}^s) \end{aligned}$$

で、表現される。

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

“PHMM based Gesture recognition”

Ken Masumitsu, Keisuke Hayata, Tetsunori Kobayashi
Faculty of Science and Engineering, Waseda University

3 シミュレーション実験

3.1 実験条件

PHMMの過渡部の表現能力を、シミュレーション実験によって評価する。区分定常信号として方形波 ($y = 20 + N(0, 1)(0 \leq x \leq \pi/2), y = -20 + N(0, 1)(\pi/2 \leq x \leq \pi)$) を、滑らかに変化する過渡部を持つ信号として余弦波 ($y = 20 \cos(x) + N(0, 1)(0 \leq x \leq \pi)$) を用いて、PHMM、HMMそれぞれを用いて学習したモデルから、方形波、余弦波の生起する尤度と、各状態毎の遷移確率・生起確率を比較する。ただし、 $N(0, 1)$ は平均0, 分散1のガウス分布である。

学習にはEMアルゴリズムを用い、評価にはビタビアルゴリズムを用いた。

3.2 実験結果

図1に方形波、余弦波がそれぞれのモデルから生起する尤度を示す。ここで、学習と尤度計算の対象が一致する時尤度が高く、一致しない時尤度が低く抑えられることが望まれる。

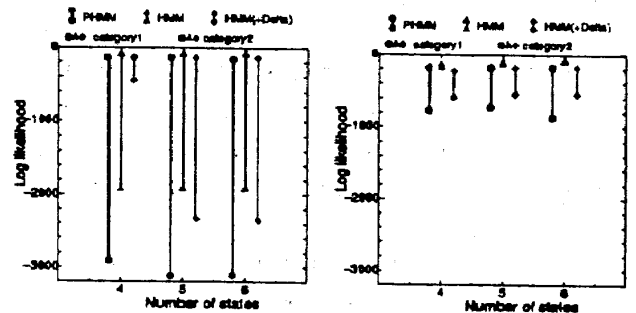


図1: 両波形の尤度の比較結果 (右:方形波による学習、左:余弦波による学習)

方形波で学習したモデルにおいては、HMM、PHMM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わかる。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るとき、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、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、学習データではない余弦波の尤度が学習データの方角波の尤度を越えてしまうという現象が生じている。

この現象を詳しく調べるために、余弦波学習時の状態数6のときにおける各時刻の遷移確率・生起確率を図2に、累積尤度変化を図3に示す。

PHMMにおいては、遷移確率に前出力依存性があるため、余弦波で学習した場合、方角波で生じるステップ上のパラメタ遷移を許さない。図2より、PHMMの状態遷移確率はt=20~22においてほぼ0になっていることがわかる。このため、PHMMは余弦波と方角波とを区別することができる。

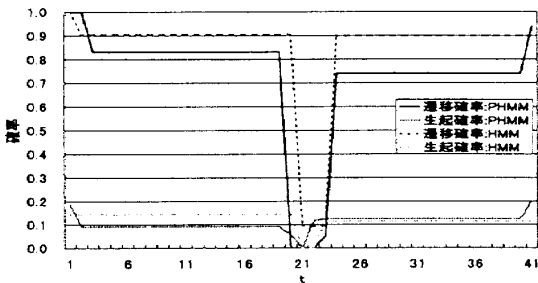


図 2: 余弦波学習-方角波入力時の遷移・生起確率変化

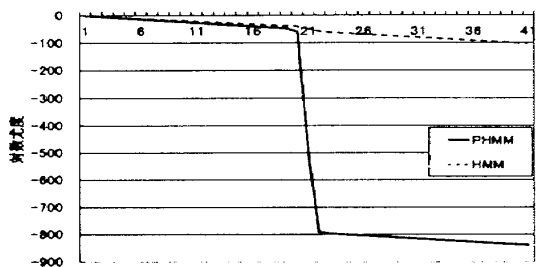


図 3: 余弦波学習-方角波入力時の累積尤度変化

4 評価実験

4.1 実験条件

認識カテゴリ：右手だけを使い表現する紛らわしいジェスチャ6語（“駄目”、“疑問”、“さよなら”、“招く”、“話す”、“好き”）

データ：40人の被験者に対し、6語を各4回演じたものをデータとして収集した。このうち、1セットをテスト用、残り3セットを学習用とし、組合せを変えて4回実験を行なう。

特徴量：各フレームにおいて、色情報と、動き情報を用いて、顔の位置と右手位置、右手面積を抽出し、そこから顔の位置に対する手の相対ベクトルを求

め、以下のような特徴量を作成する。

$$z_t - 5D = (x_t^r, y_t^r, \Delta x_t^r, \Delta y_t^r, \Delta s_t^r)$$

ただし、 (x_t^r, y_t^r) は右手の相対ベクトル、 s_t^r は右手の面積、 Δ は前時刻との差分を表す。

確率モデルの構造：F状態に関しては、先端と終端に自己ループのない、飛び越し無しの left-to-right モデル。S状態に関しては、特徴量に応じた5次元の連続空間。確率分布関数は、単一正規分布を使用。また、学習にはEMアルゴリズムを用い、認識にはビタビアルゴリズムを用いた。

比較項目：特徴量として、 $z_t - 5D$ の2種類を用いて、F状態の状態数を変えながら(7から13)、必要な状態数について調べる。これを同等の条件のHMM(HTK2.0を使用)と比較する。

4.2 実験結果

実験結果を図4に示す。

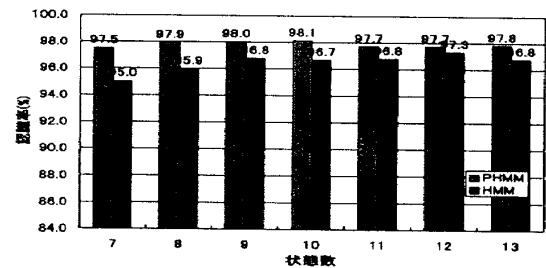


図 4: 6手話認識結果 (特徴量: $z_t - 5D$)

最良の認識結果はPHMM(状態数10)での98.1%でこれはHMMでの最良の結果であるHMM(状態数12)よりも0.8ポイント上回っている。これは30%の誤認識率改善に当たる。

以上のことから、PHMMは少ない状態数で、複雑な現象を表現できることがわかる。

5 まとめ

シミュレーション実験により、渡りの異なる二つの波形の区別において、PHMMはHMM対し優位であることが確認できた。また、ジェスチャ認識においても誤認識率を30%改善することができた。

参考文献

- [1] L. Deng, et al, IEEE Trans. on Speech and Audio Process, vol.2 no.4, pp.507-520, Sep. 1994.
- [2] C.J. Wellekens, Proc. ICASSP87, pp.383-386, 1987.
- [3] 益満 健, 古山 純子, 小林 哲則, インタラクシオン '98, pp. 13-20, March, 1998