

ガウス過程の隠れセミマルコフモデルに基づく身体動作の分節化

Body Motion Segmentation Based on Gaussian Process Hidden Markov Model

中村 友昭^{*1} アッタミム ムハンマド^{*2} 長井 隆行^{*1} 持橋 大地^{*3} 小林 一郎^{*4}
 Tomoaki Nakamura Muhammad Attamimi Takayuki Nagai Daichi Mochihashi Ichiro Kobayashi

麻生 英樹^{*5} 金子 正秀^{*1}
 Hideki Asoh Masahide Kaneko

^{*1}電気通信大学 ^{*2}玉川大学 ^{*3}統計数理研究所
 The University of Electro-Communications Tamagawa University The Institute of Statistical Mathematics

^{*4}お茶の水女子大学 ^{*5}産業技術総合研究所
 Ochanomizu University National Institute of Advanced Industrial Science and Technology

In this paper, we propose a model to segment time series data using hidden semi-Markov model (HSMM) whose emission distribution is Gaussian process (GP). The proposed model is a generative model of the time series data, which is a sequence of segments generated by GP. A length and a class of each segment are latent variables in our model, and these latent variables are estimated using Forward filtering-Backward sampling. In the experiment, we segmented motion capture data for evaluation, and the results showed that our proposed model can segment it into basic motions.

1. はじめに

人は知覚した連続的な情報を意味をもつ単位に分節化することで認知している。例えば、連続している音声波形を分節化することで意味を持つ単位である単語を認識することができ、また人の連続的な動作を分節化することで意味のある単位動作を認識することができる。人は、無数の可能性がある中で、明示的な分節点を与えられることなく学習し、情報を分節化している。ロボットにおいても、言語や動作・ジェスチャ等を柔軟に学習するために、連続的な時系列データから単位系列を教師なしで分節化する能力は非常に重要であると言える。

このような系列データを教師なしで分節化する手法として、教師なし形態素解析 [Goldwater 06][Mochihashi 09][Uchiumi 15] がある。文献 [Goldwater 06] では、階層ディリクレ過程に基づく bigram 言語モデルをギブスサンプリングで学習することで、切れ目のない文字列を単語へと分割している [Goldwater 06]。文献 [Mochihashi 09] では、階層 Pitman-Yor 過程に基づく ngram 言語モデルを Forward filtering-Backward sampling アルゴリズムを用いて学習することで、文献 [Goldwater 06] に比べ高効率かつ高精度に単語へ分割する手法 Nested Pitman-Yor Language Model (NPYLM) を提案している。さらにこれを拡張することで、文献 [Uchiumi 15] では Forward filtering-Backward sampling を適用する際に、単語の切れ目だけでなく品詞も同時にサンプリングすることで、品詞の推定も可能としている。しかし、これらの研究では、文字列のようにあらかじめ記号化されたデータを分節化することが目的であり、人が動作している際の関節の角度や位置といった連続値を取りうる系列データの分節化は考えられていない。

文献 [Taniguchi 11] では、連続値である動作系列を Infinite Hidden Markov Model [Fox 07] を用いて離散的な値にし、その離散値を NPYLM で分節化することで、単位動作を学習す

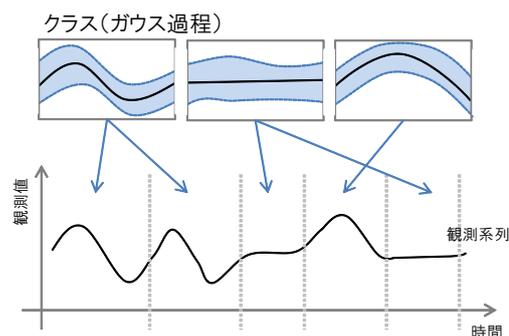


図 1: 提案手法の概要

る手法を提案している。この手法では学習したモデルを用いて動作を認識することができるが、モデル内で系列は離散化されてしまっているため単純には動作系列を生成することができない。また、ノイズなどの影響により離散化する処理で誤りが生じると分節化にも影響するといった問題がある。

そこで、本稿では系列データを離散化することなく分節化し、単位系列ごとに分類する手法を提案する。提案手法は、隠れセミマルコフモデルにおける出力をガウス過程とすることで、1つの状態が1つの連続的な単位系列を表現するモデルとなっている。図1のように、各クラスがガウス過程で表現されており、観測される系列は各クラスから生成された単位系列をつなぎ合わせることで生成される。観測系列のみからモデルを学習することで、教師なしで分節点と各单位系列のクラスを推定することができる。モデルの学習には文献 [Uchiumi 15] で提案された Forward filtering-Backward sampling を用い、単位系列の切れ目とクラスを同時にサンプリングすることで学習を行う。

連絡先: 中村 友昭, 電気通信大学知能機械工学専攻, 東京都調布市調布ヶ丘 1-5-1, 042-443-5215, tnakamura@uec.ac.jp

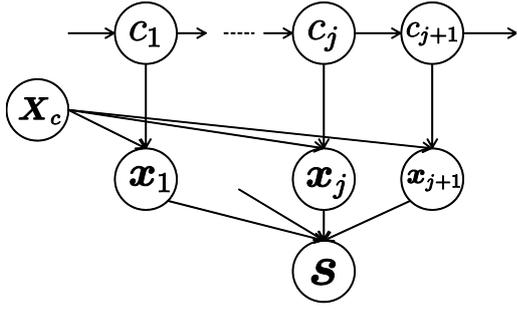


図 2: グラフィカルモデル

2. 提案手法

図 2 が提案手法のグラフィカルモデルである。図において $c_j (j = 1, 2, \dots, J)$ が単位系列のクラスを表しており、クラスと対応した \mathbf{X}_c をパラメータとするガウス過程から単位系列 \mathbf{x}_j が生成される。

$$c_j \sim P(c|c_{j-1}) \quad (1)$$

$$\mathbf{x}_j \sim \mathcal{GP}(\mathbf{x}|\mathbf{X}_{c_j}) \quad (2)$$

ただし、 \mathbf{X}_c はクラス c に分類された単位系列の集合である。これらの単位系列を結合することで、観測系列 $\mathbf{S} = \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_J$ となる。

2.1 ガウス過程

本稿では、単位系列のタイムステップ i における出力値 x_i をガウス過程回帰で学習することで、連続的な軌道として表現する。ガウス過程では、同一クラスに属する単位系列のタイムステップ i における出力 x の組 (\mathbf{i}, \mathbf{x}) が得られた時、タイムステップ i^{new} における出力 x^{new} の予測分布は次式のガウス分布となる。

$$p(x^{new} | i^{new}, \mathbf{x}, \mathbf{i}) \propto \mathcal{N}(\mathbf{k}^T \mathbf{C}^{-1} \mathbf{i}, c - \mathbf{k}^T \mathbf{C}^{-1} \mathbf{k}) \quad (3)$$

ただし、 $k(\cdot, \cdot)$ をカーネル関数として、 \mathbf{C} は、

$$C(i_p, i_q) = k(i_p, i_q) + \beta^{-1} \delta_{nm} \quad (4)$$

を要素を持つ行列であり、 β は観測値に含まれるノイズの精度を表すハイパーパラメータである。また、 \mathbf{k} は $k(i_p, i^{new})$ を要素を持つベクトルであり、 c は $k(i^{new}, i^{new})$ となるスカラーである。ガウス過程では、カーネルを用いることで複雑に変化する系列データであっても学習することができる。ここでは、ガウス過程回帰に広く使われている、以下のようなガウスカーネルを用いる。

$$k(i_p, i_q) = \theta_0 \exp\left(-\frac{1}{2} \|i_p - i_q\|^2 + \theta_2 + \theta_3 i_p i_q\right) \quad (5)$$

θ_* はカーネルのパラメータである。

さらに出力値が多次元のベクトルの場合、各次元が独立に生成されると仮定し、時刻 t の観測値 \mathbf{x}_t がクラス c に対応するガウス過程から生成される確率 $\mathcal{GP}(\mathbf{x}_t | \mathbf{X}_c)$ を以下のように計算する。

$$\begin{aligned} \mathcal{GP}(\mathbf{x}_t | \mathbf{X}_c) &= p(x_{t,0} | i = 0, \mathbf{X}_c, \mathbf{I}_t) \\ &\quad \times p(x_{t,1} | i = 1, \mathbf{X}_c, \mathbf{I}_t) \\ &\quad \times p(x_{t,2} | i = 2, \mathbf{X}_c, \mathbf{I}_t) \cdots \end{aligned} \quad (6)$$

Algorithm 1 Blocked Gibbs Sampler

```

1: 以下を収束するまで繰り返す
2: for  $n = 1$  to  $N$  do
3:   for  $j = 1$  to  $J_n$  do
4:      $N_{c_{nj}} - = 1$ 
5:      $N_{c_{nj}, c_{n,j+1}} - = 1$ 
6:     if  $j \neq 0$  then
7:       単位系列  $\mathbf{x}_{nj}$  を  $\mathbf{X}_{c_{nj}}$  から削除
8:     end if
9:   end for
10:
11:   // 単位系列とクラスをサンプリング
12:    $\mathbf{x}_{n*}, c_{n*} \sim P(\mathbf{x} | \mathbf{s}_n)$ 
13:
14:   for  $j = 1$  to  $J_n$  do
15:      $N_{c_{nj}} ++$ 
16:      $N_{c_{nj}, c_{n,j+1}} ++$ 
17:     if  $j \neq 0$  then
18:       単位系列  $\mathbf{x}_{nj}$  を  $\mathbf{X}_{c_{nj}}$  に追加
19:     end if
20:   end for
21: end for

```

この確率を用いることで、似た単位系列を同じクラスへ分類することが可能となる。

2.2 学習

2.2.1 Blocked Gibbs Sampler

観測系列内の単位系列とそのクラスを動的計画法によりサンプリングすることで推定する。動的計画法を効率的に計算するために、1つの観測系列内の単位系列とそのクラスをまとめてサンプリングする Blocked Gibbs Sampler を用いる。まず、全ての観測系列をランダムに分節化、分類することで初期化する。次に、ある観測系列 \mathbf{s}_n を分節化して得られた全ての単位系列 $\mathbf{x}_{nj} (j = 1, 2, \dots, J_n)$ を、それらが分類されているクラスからすべて削除し、各クラスのガウス過程のパラメータ \mathbf{X}_c と遷移確率 $P(c|c')$ を更新する。次に、後に示す Forward filtering-Backward sampling により、新たな単位系列 $\mathbf{x}_{nj} (j = 1, 2, \dots, J_n)$ とそれらのクラス $c_{nj} (j = 1, 2, \dots, J_n)$ を推定し、各クラスと対応したガウス過程のパラメータ \mathbf{X}_c と遷移確率 $P(c|c')$ を更新する。以上の操作を全ての系列に対して繰り返すことで、観測系列の分節点とガウス過程のパラメータを交互に最適化することができる。Algorithm 1 が Blocked Gibbs Sampler のアルゴリズムである。ただし、 $N_{c_{nj}}, N_{c_{nj}, c_{n,j+1}}$ は遷移確率 (式 (8)) を計算するためのパラメータである。

2.2.2 Forward filtering-Backward sampling

本稿では文献 [Uchiumi 15] と同様に、単位系列とそのクラスの両方を隠れ変数とみなして、Forward filtering-Backward sampling により同時にサンプリングする。

まず、Forward filtering では、ある観測系列 \mathbf{s} のタイムステップ t を終端とする k サンプル $\mathbf{s}_{t-k:k}$ が単位系列であり、そのクラスが c である確率を次のように計算する。

$$\begin{aligned} \alpha[t][k][c] &= \mathcal{GP}(\mathbf{s}_{t-k:k} | \mathbf{X}_c) \\ &\quad \times \sum_{k'=1}^K \sum_{c'=0}^C p(c|c') \alpha[t-k][k'][c'] \end{aligned} \quad (7)$$

C はクラス数、 K は単位系列の最大の長さとなる。 $p(c|c')$ はクラスの遷移確率を表しており、次式のようになる。

$$p(c|c') = \frac{N_{c'c} + \alpha}{N_{c'} + C\alpha} \quad (8)$$

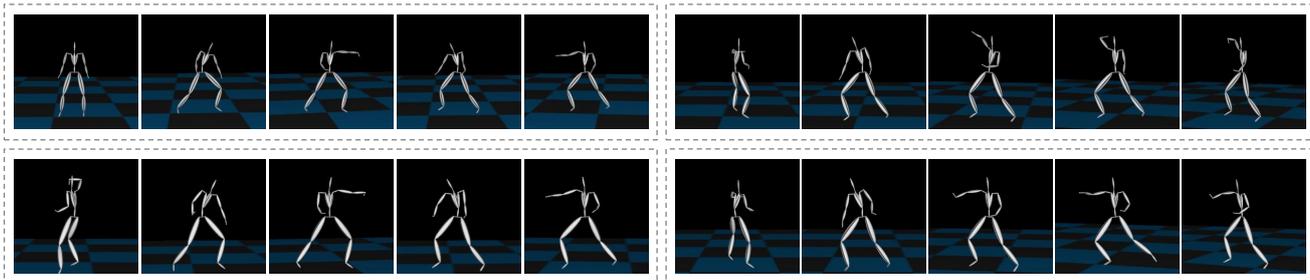


図 3: 実験に用いた動作データ

Algorithm 2 Forward filtering-Backward sampling

```

1: // Forward filtering
2: for  $t = 1$  to  $T$  do
3:   for  $k = 1$  to  $K$  do
4:     for  $c = 1$  to  $C$  do
5:        $\alpha[t][k][c]$  を計算
6:     end for
7:   end for
8: end for
9:
10: // Backward sampling
11:  $t = T, j = 1$ 
12: while  $t > 0$  do
13:    $k, c \sim p(\mathbf{x}_j | \mathbf{s}_{t-k:t}) \alpha[t][k][c]$ 
14:    $\mathbf{x}_j = \mathbf{s}_{t-k:t}$ 
15:    $c_j = c$ 
16:    $t = t - k$ 
17:    $j = j + 1$ 
18: end while
19: return  $(\mathbf{x}_{J_n}, \mathbf{x}_{J_n-1}, \dots, \mathbf{x}_1), (c_{J_n}, c_{J_n-1}, \dots, c_1)$ 

```

ただし、 $N_{c'}$ と $N_{c'c}$ は、それぞれクラスが c' となった単位系列の数と、クラスが c' から c に遷移した回数を表している。 k' と c' は、 $\mathbf{x}_{t-k:k}$ の前の単位系列の長さでクラスを表しており、式 (7) では全ての長さでクラスにおいて周辺化されている。また、 $t-k < 0$ となる場合 $\alpha[t][k][*] = 0$ であり、 $\alpha[0][0][*] = 1.0$ である。式 (7) は漸化式になっており、 $\alpha[1][1][*]$ から計算をすることで、全てのパターンを動的計画法により計算することができる。

次に、前向き確率に従い、後ろ向きに単位系列の長さでクラスをサンプリングすることで、観測系列の分節化とクラスを決めることができる。以上のアルゴリズムを、Algorithm 2 に示す。

3. 実験

提案手法の有効性を検証するために、実際のモーションキャプチャデータから身体動作の分節化を行った。データには、Mocapdata.com^{*1} で公開されている、空手の型のデータ^{*2} (図 3) を用いた。空手の型には、パンチや防御といった定型動作が複数回出現し、正解を決めることが容易であるため型のデータを用いた。使用した型は平安初段と呼ばれる、型の中では最も基本的な型であり、パンチ (追い突き)・下段の防御 (下段払い)・上段の防御 (上げ受け) から構成されている。

*1 <http://www.mocapdata.com/>

*2 http://mocapdata.com/product.cgi?product_id=10019

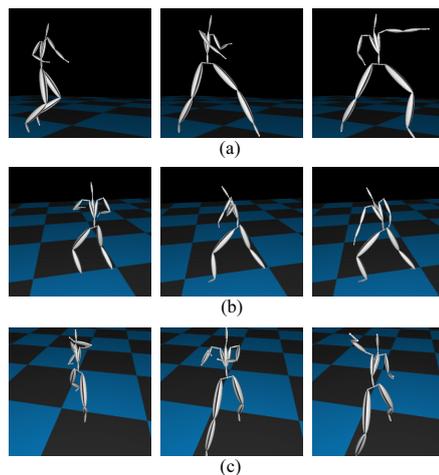


図 4: 型に含まれる基本動作の例 (a) 左パンチ (左追い突き), (b) 左下段ガード (左下段払い), (c) 右上段ガード (右上げ受け)

この型で使用されている基本動作の一部が図 4 である。本実験では、この型の前半部分をブロックドギブスサンプラーを適用するため 4 分割し、4 つの時系列データとして用いた。それぞれの動作に含まれる基本動作は以下のとおりである。

1. 左下段ガード, 右パンチ, 右下段ガード, 左パンチ
2. 左下段ガード, 右上段ガード, 左上段ガード, 右上段ガード
3. 左下段ガード, 右パンチ, 右下段ガード, 左パンチ
4. 左下段ガード, 右パンチ, 左パンチ, 右パンチ

モーションキャプチャの動作を、腰が正面かつその座標が $(0,0,0)$ となるよう正規化した右手と左手の正面から見た際の 2 次元の座標を用いた。実験では比較のため、HDP-HMM による分節化 [Beal 01] と、HDP-HMM によって離散化された状態に対して NPYLM [Mochihashi 09] を用いて分節化する手法と比較した。

図 5 が分節化の結果である。横軸が各データのフレーム番号であり、色が各フレームが分類された動作クラスを表している。この結果より、HDP-HMM では正解に比べて短い分節を推定している。これは、HDP-HMM における出力分布であるガウス分布では、動作の軌跡を表現することができなかった

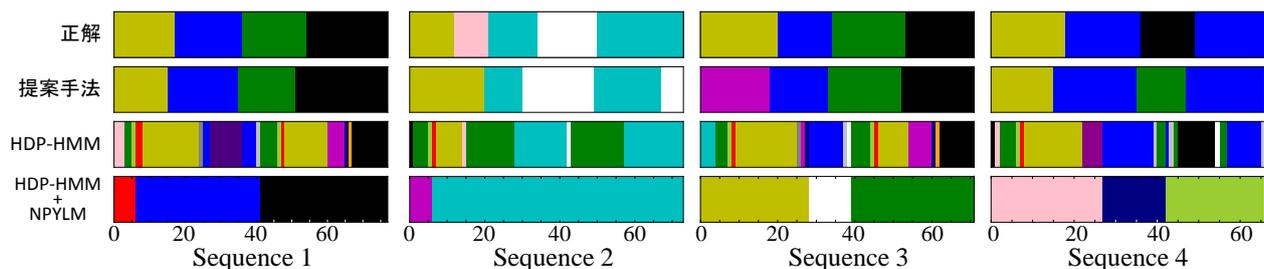


図 5: 各手法における動作の分節, 分類結果

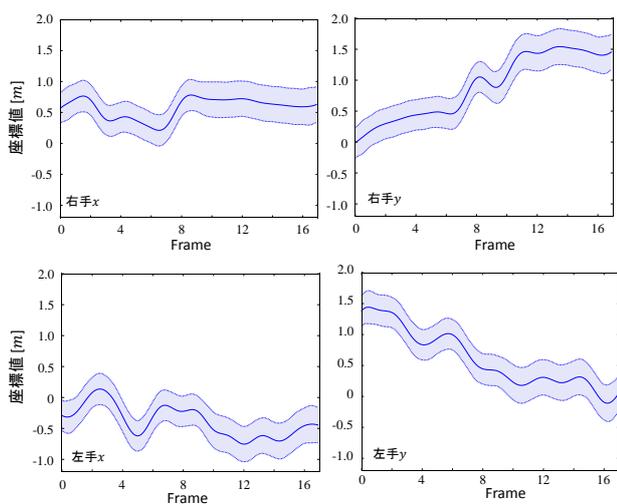


図 6: 右上段ガードの動作として学習されたガウス過程

ことが原因として挙げられる。さらに、この HDP-HMM によって離散化された状態に対して、NPYLM による分節化した結果では、正解に比べて大きな分節を推定しており、定型的な動作パターンを抽出することができていない。これは、HDP-HMM によって 1 フレームのみの分節が多く出現しており、これがノイズとなり離散化された状態に定型的なパターンが現れなかったことが原因であると考えられる。一方、提案手法では、他の手法に比べて正解に近い分節・分類ができています。図 6 が、実際に図 4(c) の右上段ガードの動作として学習されたガウス過程である。x 座標が正面から見た際の手の左右の動き、y 座標が正面から見た際の手の上下の動きである。図 4(c) のように、左手を高い位置から脇に引き、それと同時に右手を上げている動作が表現できている。

4. まとめ

本稿では、ガウス過程を出力分布とした隠れセミマルコフモデルを用いることで、動作データを分節化、分類する手法を提案した。提案手法では、動作の切れ目と、動作のクラスを Forward filtering-Backward sampling により同時に推定することにより、動作の分節化を行った。実験により、他の手法に比べて正しい分節を推定することができた。しかし提案手法では、動作を分類するクラス数を与える必要があり、他の手法との公平な比較とはなっていない。そこで、今後、ディリクレ過程の導入によるクラス数の自動推定を行う予定である。ま

た、今回用いた特徴量は、両手の xy 座標のみであったが、今後全身の特徴量を使用することで、より複雑な動作を扱うことを考えている。

謝辞

本研究は、JSPS 科研費 26280096 の助成を受けて実施したものである。

参考文献

- [Beal 01] Beal, M. J., Ghahramani, Z., and Rasmussen, C. E.: The infinite hidden Markov model, in *Advances in neural information processing systems*, pp. 577–584 (2001)
- [Fox 07] Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S.: The sticky HDP-HMM: Bayesian non-parametric hidden Markov models with persistent states, *Arxiv preprint* (2007)
- [Goldwater 06] Goldwater, S.: *Nonparametric Bayesian Models of Lexical Acquisition*, PhD thesis, Brown University (2006)
- [Mochihashi 09] Mochihashi, D., Yamada, T., and Ueda, N.: Bayesian Unsupervised Word Segmentation with Nested Pitman-Yor Language Modeling, in *ACL-IJCNLP*, Vol. 1, pp. 100–108 (2009)
- [Taniguchi 11] Taniguchi, T. and Nagasaka, S.: Double articulation analyzer for unsegmented human motion using pitman-yor language model and infinite hidden markov model, in *IEEE/SICE International Symposium on System Integration*, pp. 250–255 (2011)
- [Uchiumi 15] Uchiumi, K., Hiroshi, T., and Mochihashi, D.: Inducing Word and Part-of-Speech with Pitman-Yor Hidden Semi-Markov Models, in *ACL-IJCNLP* (2015)