

階層 Pitman-Yor 言語モデルを用いた動作解析

Motion Segmentation with Hierarchical Pitman-Yor Language Model

長坂翔吾*1 谷口忠大*1
Shogo Nagasaka Tadahiro Taniguchi

*1 立命館大学
Ritsumeikan University

This paper presents a motion segmentation method based on Sticky Hierarchical Dirichlet Process-Hidden Markov Model (Sticky HDP-HMM) and Hierarchical Pitman-Yor Language Model (HPYLM). Sticky HDP-HMM enables us to code motion data into symbols. HPYLM is one of the best smoothing method and Nested Pitman-Yor Language Model that is based on HPYLM is used for unsupervised morphological analyzer. These methods enable us to segment motion although we have only unsegmented data. In this paper, In order to verify effectiveness of these methods, we make experiment in segmentation for data which is measured by motion capture and compare our method with Minimum Description Length method which is also used for motion segmentation.

1. はじめに

幼児は親の動作を観察し模倣することにより様々な動作の獲得を行っている。このような日常生活での教示者とのインタラクションによる模倣学習を行うことができれば、教示者の観察と模倣により環境に適応した動作を獲得できるロボットが制作できると考えられる。

近年のモーションキャプチャ技術の向上により人間の身体位置、角度の測定精度が向上し、教示者の動作の観察、分析が容易になった。しかしモーションキャプチャで収集できる情報は連続時系列データである。人間であれば連続時系列データより無意識に動作単位を抽出することができるが、ロボットでは連続時系列データから動作の境界を区別することができないため、どの動作について模倣学習を行うのかを判断することができない。よって、多くの研究では特定の動作のみのデータを入力する方法が用いられてきた [稲邑 04]。また鯨島らのモデル [鯨島 02] では動作の分節はダイナミクスの変移とし、動作単位への分割を行っている。しかし、これらの手法によって分節化された動作は断片的であり、人間が認識する動作の概念とは離れている。N-gram モデルによる符号データの分節化手法としては最小記述長基準 (Minimum Description Length, MDL) を用いた谷口の手法 [谷口 09] が挙げられる。このモデルではダイナミクスの変化による符号化から、さらに、符号列の分節化を行うことで人間が認識する動作単位に近い動作の抽出を行っている。しかし、符号符号列の分節化は改善されたが、この手法では MDL の計算量を最適化するために出現頻度が低い単語についてはスコア計算を打ち切ってしまうため、局所解に収束してしまう可能性がある。また、[Taniguchi] では、ギブスサンプリングによる MDL によって大域解の探索が可能になったが、この手法では動作の分割により新たな単語が出現するたびに辞書の生成と単語分割のスコアの更新を繰り返すため、大量のデータを処理することが難しい。

これらの問題点を解決するために、提案手法では階層 Pitman-Yor 言語モデル (Hierarchical Pitman-Yor Language Model, HPYLM) を拡張した言語モデルである Nested Pitman-Yor Language Model (NPYLM) を用いた教師なし形

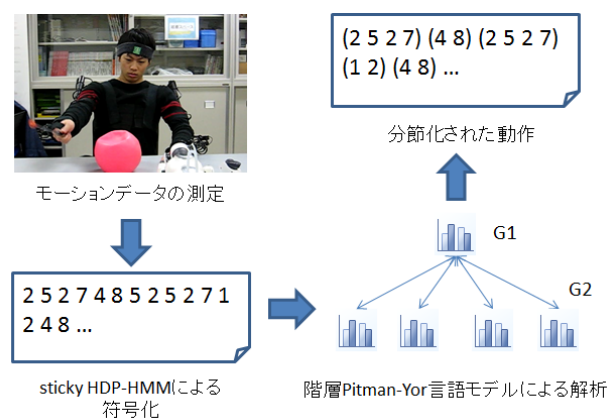


図 1: 提案手法概要

態素解析 [持橋 09] を利用する。NPYLM では、HPYLM に単語の N-gram モデルを組み込むことにより辞書データなしでの形態素解析を実現している。提案手法では、文字を符号、動作を単語に置き換えることでこの教師なし形態素解析の手法を適用し、その有用性について検証する。

2. 提案手法

提案手法ではモーションキャプチャ等により計測した動作データを動作単位ごとに分節化することを目的とする。提案手法の全体的な処理の流れを図 1 に示す。以下、各手順について説明する。

2.1 Sticky HDP-HMM による動作の符号化

モーションキャプチャで得られる情報は身体位置や身体角といった高次元の時系列データであり、形態素解析の手法を用いるには符号化が必要である。

連続時系列データを符号化するには広く Hidden Markov Model が用いられる [Bishop 06]。しかし、通常の HMM では、状態数を事前に決定しないとイケないという問題がある。これに対し、近年、ディリクレ過程 (Dirichlet Process, DP) により隠れ状態数を柔軟に決定するノンパラメトリックベイズモデルによるアプローチが注目されている [Teh 06b]。ノン

パラメトリックベイズモデルに基づき HMM の状態数を自動決定出来るようにしたのが Beal らの Infinite HMM である [Beal 02]. しかし, Infinite HMM には計算量や自己遷移確率を高めるバイアスを理論的に加えにくいといった問題点があった. Fox らは自己遷移確率のバイアスを理論的に加え, また, 計算速度も Gibbs Sampling のブロック化により速められる Sticky HDP-HMM を提案した [Fox 08]. 濱畑らはこれを用いて動作のモデル化, 符号化を行っている [濱畑 10]. 本研究でもこれに倣い, Sticky HDP-HMM を動作の符号化に用いる. Sticky HDP-HMM については [Fox 08] で詳しく解説されているため, ここでは説明を省略する.

2.2 教師なし形態素解析による分節化

Sticky HDP-HMM により符号化されたデータを HPYLM による教師なし形態素解析により動作の符号列の分節化 (チャンク化) を行う.

[持橋 09] での形態素解析では入力として複数の文が必要となるが, 今回の提案手法では多様な動作を含む動きをモーションキャプチャにより複数回計測し, 各計測データを Sticky HDP-HMM により符号化した符号列を文として入力する. これにより, 動作の符号列を分節化し動作単位を抽出することができる.

3. HPYLM による形態素解析

動作の分節化には, 自然言語における単語のような動作単位が存在しないため, 入力データより基準となる動作単位を抽出しなければならない. また, 従来の形態素解析では未知語に対する N-gram 確率の計算が適切に行えないため, 未知語が含まれる文書ではパフォーマンスが低下するという問題点が挙げられる. 教師なし形態素解析では解析対象となる単語が全て未知語になるため, この問題を解消する必要がある.

[持橋 09] での形態素解析手法では単語 N-gram モデルに加え文字 N-gram モデルを利用し, 各 N-gram モデルに Pitman-Yor 過程によるスムージングを行うことで未知語や低頻度語に対する頑強性を向上させている. また, この言語モデルを利用し動的計画法での単語サンプリングを繰り返すことにより, 入力データのみからの単語分割を行っている.

以下に階層 Pitman-Yor 言語モデルによる形態素解析について説明する.

3.1 階層 Pitman-Yor 過程

HPYLM は階層 Pitman-Yor 過程を利用した N-gram 言語モデルである. 階層 Pitman-Yor 過程は Dirichlet 過程の拡張である Pitman-Yor 過程を階層化しそれぞれの基底測度に Pitman-Yor 過程をもつ確率過程である.

階層 Pitman-Yor 過程では文脈 $h = w_{t-n} \dots w_{t-1}$ に続く単語 w の確率は

$$p(w|h) = \frac{c(w|h) - d \cdot t_{hw}}{\theta + c(h)} + \frac{\theta + d \cdot t_h}{\theta + c(h)} p(w|h') \quad (1)$$

として表される. h' は h より一つオーダーを落した文脈であり $h' = w_{t-n-1} \dots w_{t-1}$ である. よって $p(w|h')$ は一つ短い文脈での単語 w の事前確率となり, 各文脈で再帰的に計算される. $c(w|h)$ は文脈 h での w の出現回数であり, $c(h)$ は文脈 h における各単語の出現回数の総和である. t_{hw} は文脈 h' から生成したと推定される回数であり, t_h は文脈 h での各単語での t_{hw} の総和である. また, d, θ は階層 Pitman-Yor 過程のパラメータであり, ギブスサンプリングを利用することにより推定することができる.

この言語モデルを用いることにより未知語に対する確率のスムージングを行うことができる. また Pitman-Yor 過程の階層化により特定の文脈では出現しない単語の確率も一つ短い文脈の階層の確率で補完することができる.

3.2 NPYLM への拡張

HPYLM では文脈となる単語が観測できたとき, 次の単語の N-gram 確率を計算することができた. しかし, 単語ユニグラムの場合では式 (1) における $p(w|h')$ は辞書が与えられている場合はその語彙数の逆数を設定すればよいが, 辞書がない場合にはその文書での全ての部分文字列が単語となる可能性があるため語彙数を設定することができない.

これを問題を解決するために, 単語ユニグラムの基底速度に文字での HPYLM を適用する (NPYLM) を利用する. 文字 N-gram を利用することにより, 単語中に出現する部分文字列の N-gram 確率を求めることができる.

3.3 ブロック化ギブスサンプリング

HPYLM の階層化により辞書なしでの N-gram 確率の計算が可能になった. この NPYLM を利用して入力文書の形態素解析を行う. NPYLM ではブロック化ギブスサンプリングと動的計画法を用いて高速に単語の分割を行っている.

ブロック化ギブスサンプリングでは文に含まれる単語分割を言語モデルから削除し, 新たな単語分割を言語モデルよりサンプリングする. これを言語モデルに加えることにより言語モデルの更新する. これを入力文書での全ての文で, ランダムな順番で繰り返すことにより, 言語モデルの最適化を行う.

NPYLM での単語の分割は HMM のベイズ学習で利用される Forward filtering-Backward sampling アルゴリズムを利用する.

まず Forward filtering では前向き確率 $\alpha[t][k]$ を導入しその計算を行う. $\alpha[t][k]$ は文 s において部分文字列 $c_1 \dots c_t$ から最後の k 文字が単語として分割される確率である. この前向き確率を以下の計算式

$$\alpha[t][k] = \sum_{j=1}^{t-k} p(c_{t-k+1}^t | c_{t-k-j+1}^{t-k}) \cdot \alpha[t-k][j] \quad (2)$$

ただし $\alpha[0][0] = 1$ とする

により計算する. 式 (2) は再帰的に計算されるため, 部分文字列以前の確率が周辺できる.

次に Backward sampling では計算された前向き確率を用いて文末からサンプリングする. 前向き確率により文中での部分文字列の確率が計算できているため, 文での部分文字列 c_{t-k+1}^t がサンプリングされる確率は

$$p(w_{last} | c_{t-k+1}^t) \cdot \alpha[t][k] \quad (3)$$

によって計算できる. ただし, w_{last} は最後に分割された部分文字列を表し初期値は文の終端記号である. 式 (3) の値に比例するような確率で分割位置 k をサンプルすることで最後の単語分割を決定することができる. この手続きを文頭まで繰り返すことにより, 文から単語をサンプリングすることができる.

4. 実験

NPYLM による動作分節化の妥当性を検証するために実験を行った. 実験ではモーションキャプチャから取得したデータからの動作の抽出を目的とする. 実験では被験者に 3 つの物体を与え, それらの物体を操作する動作をモーションキャプ

```

10 4 16 4 6 11 6 16 4 6 11 6 16 17 4 16 4 10 4 10 4 10
10 4 10 4 10 4 16 4 6 11 6 16 4 6 11 6 16 4 6 11 6 16 4 10
10 4 16 4 6 11 6 16 4 6 11 6 16 17 10 4 10 4 16 4 6 11 6 16 17
4 16 4 10
    
```

図 2: Sticky HDP-HMM からの出力符号列

```

(10 4) (16 4 6 11 6) (16 4 6 11 6) (16 17) (4 16 4) (10 4) (10
4) (10)
(10 4 10 4) (10 4) (16 4 6 11 6) (16 4 6 11 6) (16 4 6 11 6)
(16) (4 16 4) (10)
(10 4) (16 4 6 11 6) (16 4 6 11 6) (16 17) (10 4 10 4) (16 4 6
11 6) (16) (17) (4 16 4) (10)
    
```

図 3: NPYLM によるからの分節化結果

チャにより計測する。計測した動作の時系列データを Sticky HDP-HMM により符号化し、符号列を NPYLM により分節化する。

実験では機械式モーションキャプチャである Meta Motion 社製の Gypsy5 Torso を用いて関節角を計測した。モーションキャプチャでは頭、首、両鎖骨、両肘、両手首、胸、腰のオイラー角の 36 次元のデータを計測することができる。また、モーションキャプチャでのフレームレートは 60[Hz] に設定した。

Sticky HDP-HMM のハイパーパラメータは $\alpha = 0.1, \gamma = 0.1, \kappa = 0.9$ に設定し、分散の事前分布である逆ウィシャート分布のハイパーパラメータ $\Delta = dI$ での d を 0.5×10^{-4} と設定した。また、NPYLM での反復回数は 200 回とし、各オードダでの Pitman-Yor 過程でのハイパーパラメータの初期値は $d = 0.5, \theta = 0.1$ と設定した。

4.1 実験 1

実験 1 では NPYLM での動作データの分節化について評価する。

4.1.1 実験条件

解析対象の動作として、被験者に操作対象物である指示棒、ボール、AIBO を与え、それらの対象物を 20 秒間で自由に選択しそれぞれの操作について決った動作で操作するよう指示した。この動作を 1 セットとして 3 セットの計測データを解析対象とした。

4.1.2 実験結果

Sticky HDP-HMM により出力された符号列を図 2 に示す。この符号列を NPYLM への入力とする。

NPYLM からの出力結果を図 3 に示す。括弧で囲まれた符号列が一つの動作単位を示す。図 3 より動作単位として (10 4), (16 4 6 11 6), (16 17), (10), (4 16 4), (10 4 10 4), (16), (17) が動作単位として抽出されたことが分る。得られた動作単位より、実際の動作データを再生したところ、動作単位 (16 4 6 11 6) は AIBO を持ち上げる動作 (図 4), (4 16 4) はボールを転がす動作に対応していた (図 5)。また、動作単位 (10 4) は (10 4 10 4) の部分動作であり、これは棒を一回振る動作に該当する (図 6)。また (16 17), (16), (17) の動作は 1 秒未満の動作であり、各操作の切り替わりを示すものと考えられる。

これらの結果より、NPYLM による動作の分節化によって連続的な動作データから三つの対象物を操作するという動作をそれぞれ区別して抽出することができたと考えられる。

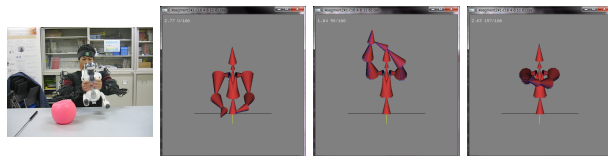


図 4: AIBO を持ち上げる動作 (16 4 6 11 6)

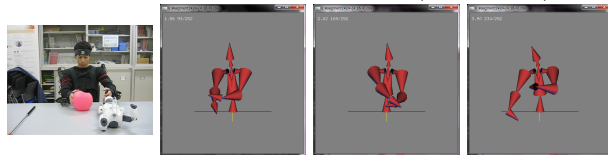


図 5: ボールを転がす動作 (4 16 4)

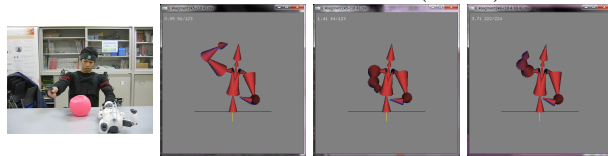


図 6: 指示棒を振る動作 (10 4 10 4)

4.2 実験 2

実験 2 では、先行研究である MDL による動作の分節化について、入力データの増加による計算時間の変化について比較する。

4.2.1 実験条件

入力データは実験 1 での 3 セット分の入力データの解析にかかった計算時間を基本計算時間とし、6, 15, 30 セットでのデータの解析時間を測定する。MDL での学習の反復回数は 200 回に設定した。

4.2.2 実験結果

図 7 より NPYLM では MDL に比べデータの増加に対し計算量の増加が緩やかであることが分る。MDL では分節化により単語と辞書が変化するため、反復ごとにほぼ情報量の再計算が必要となり、これはデータ数が増加するほど計算時間が増加すると考えられる。一方で、NPYLM では計算時間は入力文書中の文に応じて決定するため大量のデータを高速に処理できると考えられる。[持橋 09] でも動的計画法によるサンプリングによって計算時間が大幅に減少したことが報告されているが、本実験においても動的計画法によるサンプリングによって計算効率が向上したことが再確認された。

5. 考察とまとめ

本研究では階層 Pitman-Yor 言語モデルを用いた教師なし学習での動作の分節化手法について提案し、その有用性について検証を行った。NPYLM を用いることで教師データとなる辞書データが存在していない場合でも入力データの N-gram モデルにより符号データからの動作単位の抽出を行うことができる。本手法により、NPYLM による連続時系列動作の分節化が効率よく行えることが示された。本手法の改良点として、動作の符号化と言語モデルの相互の連携が挙げられる。これにより、言語モデルでの単語のつながりを動作の符号化に紐込むことができ、符号化のパフォーマンスの向上が期待できる。また、符号化に用いている Sticky HDP-HMM はデータの増加によって計算量が大きく増加するため、大量のデータを効率的に解析するためには符号化手法について改善する必要がある。

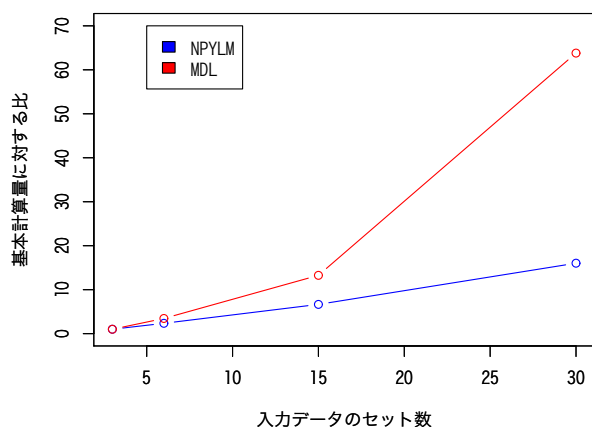


図 7: 計算時間の比較

参考文献

- [Beal 02] Beal, M., Ghahramani, Z., and Rasmussen, C.: The infinite hidden Markov model, *Advances in Neural Information Processing Systems*, Vol. 1, pp. 577–584 (2002)
- [Bishop 06] Bishop, C. M.: *Pattern Recognition And Machine Learning (Information Science and Statistics)*, Springer-Verlag, new edition edition (2006)
- [Fox 08] Fox, E., Sudderth, E., Jordan, M., and Willsky, A.: The sticky HDP-HMM for systems with state persistence, in *Proceedings of the International Conference on Machine Learning* (2008)
- [Taniguchi] Taniguchi, T., Hamahata, K., and Iwahashi, N.: Imitation learning architecture for unsegmented human motion using sticky HDP-HMM and MDL-based phrase extraction method
- [Teh 06a] Teh, Y. W.: A hierarchical Bayesian language model based on Pitman-Yor processes, in *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, ACL-44, pp. 985–992, Stroudsburg, PA, USA (2006)
- [Teh 06b] Teh, Y., Jordan, M., Beal, M., and Blei, D.: Hierarchical dirichlet processes, *Journal of the American Statistical Association*, Vol. 101, No. 476, pp. 1566–1581 (2006)
- [稲邑 04] 稲邑 哲也, 中村 仁彦, 戸嶋 巖樹, 江崎 英明: ミメシス理論に基づく見まね学習とシンボル創発の統合モデル, *日本ロボット学会誌*, Vol. 22, No. 2, pp. 256–263 (2004)
- [高野 05] 高野 渉, 中村 仁彦: 統計的相関に基づく動作パターンの分節化, *人工知能学会全国大会予稿集*, 3F1-02 (2005)

[鮫島 02] 鮫島 和行, 片桐 憲一, 銅谷 賢治, 川人 光男: モジュール競合による運動パターンのシンボル化と見まね学習, *電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理*, Vol. 85, No. 1, pp. 90–100 (2002)

[持橋 09] 持橋 大地, 山田 武士, 上田 修功: バイズ階層言語モデルによる教師なし形態素解析, *情報処理学会研究報告* (2009)

[谷口 09] 谷口 忠大, 岩橋 直人: 複数予測モデル遷移の N-gram 統計に基づく非分節運動系列からの模倣学習手法, *知能と情報*, Vol. 21, No. 6, pp. 1143–1154 (2009)

[濱畑 10] 濱畑 慶太, 谷口 忠大, 岩橋 直人, 西川 郁子: ディリクレ過程と相互情報量による非分節対象物操作からの動作抽出, *人工知能学会全国大会論文集*, 第 24 巻 (2010)