

# ノンパラメトリックベイズ二重分節解析器に関する研究

## A Preliminary Study of Nonparametric Bayesian Double Articulation Analyzer

長坂 翔吾      谷口 忠大  
Nagasaka Shogo      Taniguchi Tadahiro

立命館大学  
Ritsumeikan University

In this paper, we report our preliminary results about nonparametric Bayesian double articulation analyzer which is a hierarchical Bayesian model for time series data. We introduce a novel generative model which has two-layer hierarchical structure. The two layers correspond to words and phonemes in speech signal. We show that the nonparametric double articulation analyzer can extract words and phonemes from sequences of simple synthetic data.

### 1. 研究目的

二重分節構造は人間の音声言語に潜む構造である。音声時系列データは単語が連なり文になり、単語は音素、もしくは音節の連なりから構成される。音素を表すラベルはある程度の時間長持続し、その間、類似性の高い特徴量を持つ音響信号を生成する。音声認識の研究においてこのような二重分節構造は広く受け入れられてきた。

一方で筆者らは、人間の動作データや、自動車運転挙動データにおいても、二重分節構造が潜んでいると考え、この構造を教師なし学習で推定する二重分節解析の手法を提案してきた [1, 2, 3]。しかし、これまでの手法では、言語モデルを前提とせず、HDP-HMM[4] や HDP-HSMM[5] といった統計手法により離散記号列に変換した後に、教師なし形態素手法 [6] を用いてチャンク化するという二段階の構成をとっていた。このような手法では離散記号列に変換する際の変換誤り、つまり、音素認識誤りに対応することができない。また、二重分節構造を持つ適切な時系列データの生成モデルを持たないため、推定結果の評価も困難であった。この問題を解決するために本稿では二重分節構造を持つ時系列データの完全な生成モデルを与える。また、導出したノンパラメトリックベイズ二重分節解析器によって簡単な人工時系列データを解析した結果について報告する。

### 2. 提案モデル

本稿では Hierarchical Dirichlet Process Hidden Semi Markov Model(HDP-HSMM)[5] を拡張することにより二重分節構造を有する時系列データを生成する生成モデルを提案する。提案モデルでは隠れ状態  $z_t$  で表わされる観測データ区間が、その状態に対応する left-to-right の隠れ状態遷移を持つ HSMM を持つと仮定しモデル化を行うことで二重の分節化を行う。本稿では上位の隠れ状態を潜在単語、潜在単語を構成する下位の隠れ状態を潜在文字とそれぞれ呼ぶ。

提案モデルのグラフィカルモデルを図 1 に示す。ここで  $\beta', \alpha', \pi'_j$  は潜在文字のパラメータであり、それぞれ潜在文字遷移の基底分布、潜在文字遷移の DP のパラメータ、各潜在文字  $j$  の遷移確率を表わす。 $\beta, \alpha, \pi_i$  はそれらに対応する潜在単語の

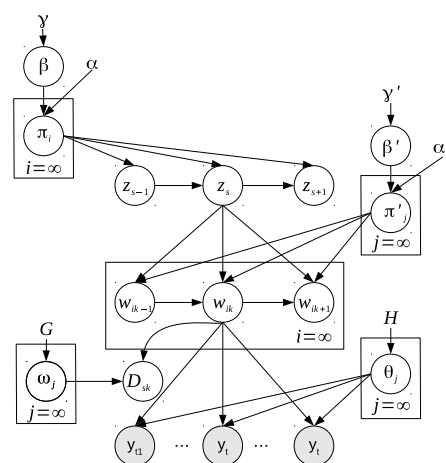


図 1: 提案モデルのグラフィカルモデル

パラメータである。 $w_{ik}$  は潜在単語  $i$  の  $k$  番目の潜在文字列を表し、一つ前の潜在文字の遷移確率よりサンプルされる。また  $D_{sk}$  は  $s$  個目の潜在単語の  $k$  番目の潜在単語の持続時間を表す。 $l_i$  は潜在単語を構成する潜在文字の長さを表わし、分布  $K$  よりサンプルされる。

HDP-HSMM では隠れ状態により決定される持続時間分布と出力分布によりデータが生成されるのに対して、提案モデルでは潜在単語  $z$  により、その潜在単語を構成する潜在文字列  $w_z$  が決まり、その後各文字列に対応する持続時間分布と出力分布より観測データが生成される。これにより潜在単語として分節化される区間は、一定の遷移パターンを持つデータとしてモデル化することができる。推定は Johnson ら [5] の提案した weak-limit approximation に基づいたブロック化ギブスサンプラーを拡張することで得る。

### 3. 実験

評価実験では二重分節構造を仮定したデータを学習データとして与え、その分節結果を評価した。

#### 3.1 実験条件

提案手法による潜在単語の推定を検証するために、二重分節構造を持つ人工データを利用した実験を行った。人工データは

連絡先: 谷口忠大, 立命館大学情報理工学部, 〒 525-8577 滋賀県草津市野路東 1-1-1, 077-561-5839, taniguchi@ci.ritsumeikan.ac.jp

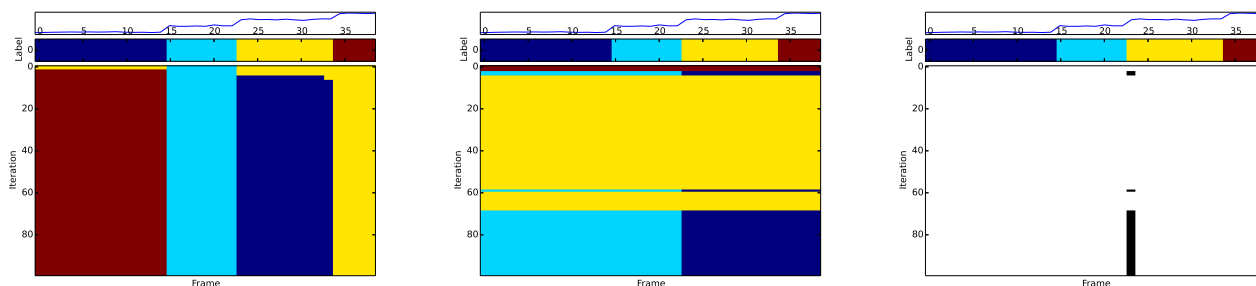


図 2: データ [0, 1], [2, 3] に関する潜在変数の推定結果 (左) 潜在文字列 (中) 潜在単語列 (右) 単語境界 .

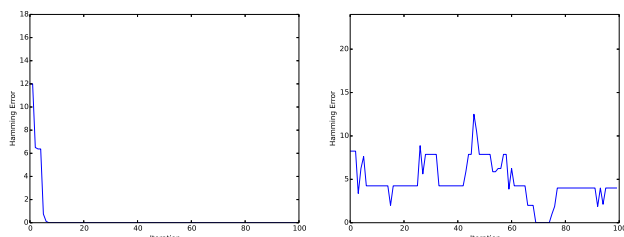


図 3: (左) 潜在文字 (右) 潜在単語の推定結果のハミング誤差

潜在文字 4 種類 (0, 1, 2, 3) からなり, 二つの潜在単語 ([0, 1], [2, 3]) からなる. 各潜在単語の持続時間分布はポアソン分布を仮定しパラメータは  $\alpha = 80, \beta = 10$  のガンマ分布からサンプルした. また出力分布は 1 次元ガウス分布を仮定し,  $\mu = 5i, \sigma^2 = 0.1$  とした ( $i$  は潜在文字の値を表す). また学習データセットは四つの潜在単語列 ([0, 1], [2, 3]), ([2, 3], [0, 1]), ([0, 1], [0, 1]), ([2, 3], [2, 3]) からそれぞれ 2 個ずつ持続時間分布, 出力分布から生成された観測データ列からなる.

潜在単語の遷移確率パラメータのハイパーパラメータは  $\gamma = 1.0, \alpha = 10.0$  に設定し最大単語数を 8 とした. 潜在単語の遷移確率パラメータのハイパーパラメータは  $\gamma' = 10.0, \alpha' = 10.0$  に設定し最大文字数を 4 とした. 単語の文字列長分布はポアソン分布を仮定しハイパーパラメータは  $\alpha = 20, \beta = 10$  に設定した. 持続時間分布のハイパーパラメータは  $\alpha = 80, \beta = 10$ , 出力分布のハイパーパラメータは  $\mu_0 = 0, \sigma_0^2 = 1.0, \kappa_0 = 0.01, \nu_0 = 1$  に設定した. またモデルの Gibbs Sampling のイテレーション回数は 100 に設定した.

### 3.2 結果

提案モデルによりサンプルされた潜在変数を図 2 に示す. 潜在単語列 ([0, 1], [2, 3]) から生成されたデータに関する推定結果である. 図中の上段は観測データのプロットであり, 中段は学習時に与えられた正解の潜在文字の境界を示す. 下段の図はサンプリング結果を示しており, 縦軸はイテレーション回数を示している.

潜在文字の推定結果と正解の潜在文字列とのハミング距離を図 3(左) に示す. 縦軸はハミング誤差, 横軸はイテレーション回数を表す. 図 3(左) よりイテレーション回数が 10 以上ではハミング誤差が 0 であることが分かる. これらの結果より提案モデルは観測データの潜在文字を正しく推定できると言える. 一方で, 図 3(右) に示すように, 潜在単語の推定結果と正解の潜在単語列とのハミング距離には誤差が残った. 提案モデルの潜在単語推定では, 同一の潜在文字列を持つ複数の単語が出現することがあり, それがハミング誤差を大きくしていた.

## 4. まとめ

本稿では二重分節構造を持つ時系列データの完全な生成モデルを与え, これに基づき導出したブロック化ギブスサンプラーを用いた人工データの解析結果を示した. より音素認識誤りが生じやすいデータや, 音声発話データなどに適用し, その有効性を示すのが今後の課題である. また, 現状では計算コストが過大になっており, これをいかに抑えるかも課題である.

## 参考文献

- [1] Tadahiro Taniguchi and Shogo Nagasaka. 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 (SII) 2011*, pp. 250–255, 2011.
- [2] Tadahiro Taniguchi, Shogo Nagasaka, Kentarou Hitomi, Nainwala P Chandrasiri, and Takashi Bando. Semiotic prediction of driving behavior using unsupervised double articulation analyzer. In *IEEE Intelligent Vehicles Symposium (IV) 2012*, pp. 849–854. IEEE, 2012.
- [3] Kazuhito Takenaka, Takashi Bando, Shogo Nagasaka, Tadahiro Taniguchi, and Kentarou Hitomi. Contextual scene segmentation of driving behavior based on double articulation analyzer. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2012*, pp. 4847–4852, 2012.
- [4] Emily B Fox, Erik B Sudderth, Michael I Jordan, and Alan S Willsky. A sticky HDP-HMM with application to speaker diarization. *The Annals of Applied Statistics*, Vol. 5, No. 2A, pp. 1020–1056, 2009.
- [5] Matthew J. Johnson and Alan S. Willsky. Bayesian non-parametric hidden semi-markov models. *Journal of Machine Learning Research*, Vol. 14, pp. 673–701, February 2013.
- [6] Daichi Mochihashi, Takeshi Yamada, and Naonori Ueda. Bayesian unsupervised word segmentation with nested pitman-yor language modeling. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp. 100–108, 2009.