

階層ディリクレ過程隠れ言語モデルへの変分ベイズ法の適用

Variational Inference for Hierarchical Dirichlet Process-Hidden Language Model

林 楓^{*1} 高野 敏明^{*2} 萩原 良信^{*2} 谷口 忠大^{*2}
 Hayashi Kaede Takano Toshiaki Hagiwara Yoshinobu Taniguchi Tadahiro

^{*1}立命館大学情報理工学研究科

Graduate School of Information Science and Engineering, Ritsumeikan University.

^{*2}立命館大学情報理工学部

Dept. of Information Science and Engineering, Ritsumeikan University.

For a computational understanding of the language acquisition by infants, researchers have developed machine learning models that can discover words from unsegmented speech signals. Nonparametric Bayes Double Articulation Analyzer (NPB-DAA) is one of the unsupervised methods for acquiring vocabulary and can simultaneously learn a language model and an acoustic model. It was shown that NPB-DAA outperformed conventional double articulation analyzer in previous studies, but performing inference with large datasets is a challenge because the inference procedure based on Gibbs sampling requires huge computational time. Markov chain Monte Carlo (MCMC) sampling is not suitable for complex bayesian models such as the generative model of the NPB-DAA called hierarchical Dirichlet process hidden language model (HDP-HLM). Variational inference provides many frameworks for approximating model posteriors efficiently. In this paper, we report our current work developing a variational inference algorithm for HDP-HLM.

1. はじめに

連続音声データからの語彙獲得は、人間の言語発達の過程で重要な役割を果たしている。人間の幼児は月齢 8 か月においても音声言語に存在する音の統計的な規則性を用いて単語を切り出せることが知られている [Saffran 96]。このような、幼児が行う単語分割への対する計算論的な理解のために、ラベル付けされていない人間の音声信号から、教師なし学習で単語を獲得するモデルを構築することは重要であると考えられる。

このような背景のもと、谷口らは人間の音声言語に含まれる二重分節構造に基づいた教師なし語彙獲得手法である NPB-DAA (Nonparametric Bayes Double Articulation Analyzer) を提案した [Taniguchi 15]。二重分節構造とは、音声学における最小単位である音素を下位層、音素で構成される単語を上位層とする二層の構造のことである。谷口らは、言語モデルと音響モデルを統合し、一つの生成モデルとして表現した階層ディリクレ過程隠れ言語モデル (HDP-HLM: Hierarchical Dirichlet Process Hidden Language Model) を提案し、これに対してブロック化ギブスサンプリングを行うことで言語モデルと音響モデルを同時に学習することを可能とした。また、階層ディリクレ過程隠れ言語モデルは HDP-HSMM (Hierarchical Dirichlet Process Hidden Semi-Markov Model) [Johnson 12] を拡張して得られる、潜在的に二重分節構造をもつ時系列データに対する生成モデルである。この研究では、言語モデルと音響モデルの同時推定により音素認識誤りに対応し、実音声データでの高精度な単語分割が可能となった。しかし、推定に膨大な計算時間がかかるため、現状では母音列のみで構成される小規模コーパスでの語彙獲得しか行われていない [Taniguchi 16]。NPB-DAA で用いている推論手法は、マルコフ連鎖モンテカルロ法 (MCMC) の一つであるブロック化ギブスサンプリングにもとづいている。一般的に、MCMC による推論は変分ベ

イズ法に比べて遅いと知られており、変分ベイズ法の利用により高速化が期待される。また、階層ディリクレ過程隠れ言語モデルを用いて、実際に人間が行うような発話からの語彙獲得を実現するには、より大規模なコーパスに対応した学習手法を検討する必要がある。

そこで、本研究では NPB-DAA の学習プロセスをマルコフ連鎖モンテカルロ法から変分ベイズ学習の枠組みに変更する。近年、データセットの大規模化に伴って、変分ベイズ学習では効率的な手法が数多く提案されている [Broderick 13, Hoffman 13]。これらを階層ディリクレ過程隠れ言語モデルに適用することで、NPB-DAA の計算時間を軽減できることが考えられる。したがって、本稿ではその第一歩として、NPB-DAA の生成モデルである階層ディリクレ過程隠れ言語モデルへの変分ベイズ法の適用を検討する。

2. 階層ディリクレ過程隠れ言語モデル

階層ディリクレ過程隠れ言語モデルは、文を構成する Language model, 単語を構成する Word model, 音の情報を持つ Acoustic model からなる、HDP-HSMM [Johnson 12] の拡張モデルである。 z_s は super state であり、潜在単語を表している。また、 i 番目の super state である $z_s = i$ は音素列 $w_i = (w_{i1}, \dots, w_{ik}, \dots, w_{iL_i})$ を持っている。ただし L_i は i 番目の単語 w_i の長さである。また、生成過程は以下のように表される。

$$\beta^{LM} \sim \text{GEM}(\gamma^{LM}) \quad (1)$$

$$\pi_i^{LM} \sim \text{DP}(\alpha^{LM}, \beta^{LM}) \quad i = 1, 2, \dots, \infty \quad (2)$$

$$\beta^{WM} \sim \text{GEM}(\gamma^{WM}) \quad (3)$$

$$\pi_j^{WM} \sim \text{DP}(\alpha^{WM}, \beta^{WM}) \quad j = 1, 2, \dots, \infty \quad (4)$$

$$w_{ik} \sim \pi_{w_{ik-1}}^{WM} \quad i = 1, 2, \dots, \infty \quad (5)$$

$$k = 1, 2, \dots, L_i \quad (6)$$

連絡先: 林 楓, 立命館大学情報理工学研究科, 〒 525-8577 滋賀県草津市野路東 1-1-1, k.hayashi@em.ci.ritsumei.ac.jp

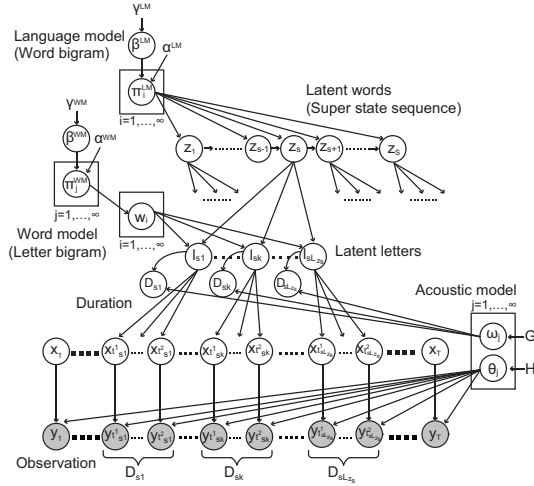


図 1: 階層ディリクレ過程隠れ言語モデルのグラフィカルモデル

$$(\theta_j, \omega_j) \sim H \times G \quad j = 1, 2, \dots, \infty \quad (7)$$

$$z_s \sim \pi_{z_{s-1}}^{LM} \quad s = 1, 2, \dots, S \quad (8)$$

$$l_{sk} = w_{z_s k} \quad s = 1, 2, \dots, S \quad (9)$$

$$k = 1, 2, \dots, L_{z_s} \quad (10)$$

$$D_{sk} \sim g(\omega_{l_{sk}}) \quad s = 1, 2, \dots, S \quad (11)$$

$$k = 1, 2, \dots, L_{z_s} \quad (12)$$

$$x_t = l_{sk} \quad t = t_{sk}^1, \dots, t_{sk}^2 \quad (13)$$

$$t_{sk}^1 = \sum_{s' < s} D_{s'} + \sum_{k' < k} D_{sk'} + 1 \quad (14)$$

$$t_{sk}^2 = t_{sk}^1 + D_{sk} - 1 \quad (15)$$

$$y_t = h(\theta_{x_t}) \quad t = 1, 2, \dots, T \quad (16)$$

ここで、GEM と DP はそれぞれ stick breaking process と Dirichlet process を表しており、上付きの文字 LM と WM はそれぞれ language model と word model を意味している。また、 β^{WM} は基底測度であり、 α^{WM} と γ^{WM} は単語を生成する word model のハイパーパラメータである。DP(α^{WM}, β^{WM}) は潜在文字 j から次の文字への遷移確率 π_j^{WM} を出力する。 β^{LM} は基底測度であり、 α^{LM} と γ^{LM} は単語バイグラムを生成するハイパーパラメータである。また、DP(α^{LM}, β^{LM}) は潜在単語 i から次の単語への遷移確率 π_i^{LM} を出力する。潜在単語 w_i に含まれる潜在文字は $\pi_{w_{ik-1}}^{WM}$ から順にサンプルされる。

HDP-HLM では、持続分布は各潜在文字 l_{sk} によって決まる。潜在文字 l_{sk} の持続時間 D_{sk} は $g(\omega_{l_{sk}})$ から生成される。ただし、 l_{sk} は s 番目の潜在単語 w_s の中の k 番目の潜在文字であり、 $\omega_{l_{sk}}$ は潜在文字 l_{sk} の持続パラメータである。また、潜在単語 w_s の持続時間は $D_s = \sum_{k=1}^{L_{z_s}} D_{sk}$ となる。

階層ディリクレ過程隠れ言語モデルでは潜在単語 z_s が潜在文字系列 $l_{sk} = w_{z_s k}$ ($k = 1, 2, \dots, L_{z_s}$) を決定する。その後 w_{z_s} に基づいて、 l_{sk} の持続時間 D_{sk} が生成され、出力分布 $h(\theta_{x_t})$ から観測データ y_t が生成される。これにより、潜在単語としてチャンク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される。

3. 階層ディリクレ過程隠れ言語モデルの変分ベイズ学習

本稿では階層ディリクレ過程隠れ言語モデルにおいて変分ベイズ学習の枠組みで推論するためのアルゴリズムについて概要を述べる。変分ベイズ法の階層ディリクレ過程隠れ言語モデルへの適用は、Beal らの隠れマルコフモデルでの変分ベイズ学習の枠組み [Beal 03] を拡張して行う。階層ディリクレ過程隠れ言語モデルでは、隠れマルコフモデルの持続時間を確率変数とした隠れセミマルコフモデルがベースになっている。さらに、二重分節構造を表現しているため、隠れ状態が z, x の 2 つになっている。変分ベイズ法の隠れセミマルコフモデルへの適用には Johnson らのアルゴリズム [Johnson 14] を用いる。Johnson らのアルゴリズムでは、隠れセミマルコフモデルの隠れ状態 x に対して message passing algorithm の forward message と backward message を計算し、各パラメータの変分事後分布を更新している。

しかし、階層ディリクレ過程隠れ言語モデルでは、隠れ状態が二階層になっているために、隠れ状態 x, z の双方を同時に message passing することは膨大なコストを要する。また、Johnson らの手法の直接的な適用では word model の推定手法を導出することが出来ない。本研究では word model の学習時には z を点推定するという近似を用いて推論手法の導出を行った。

4. まとめ

本稿では階層ディリクレ過程隠れ言語モデルと変分ベイズの枠組みでの学習法について概要を説明した。階層ディリクレ過程隠れ言語モデルは HDP-HSMM の拡張だが、二重分節構造を持つという特徴のため、局所潜在変数が 2 つあるという複雑さを持っている。提案手法では、これに対して二段階の message passing を考えた上で適切に近似することで、変分ベイズの枠組みでの学習手法を導出した。結果の詳細には口頭発表にて述べる。

参考文献

[Saffran 96] Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport. "Statistical learning by 8-month-old infants." *Science* 274.5294 (1996): 1926-1928.

[Johnson 12] Johnson, Matthew J., and Alan Willsky. "The hierarchical Dirichlet process hidden semi-Markov model." *arXiv preprint arXiv:1203.3485* (2012).

[Taniguchi 15] Taniguchi, Tadahiro, Ryo Nakashima, and Shogo Nagasaka. "Nonparametric Bayesian Double Articulation Analyzer for Direct Language Acquisition from Continuous Speech Signals." *arXiv preprint arXiv:1506.06646* (2015).

[Beal 03] Beal, Matthew James. *Variational algorithms for approximate Bayesian inference*. London: University of London, 2003.

[Broderick 13] Broderick, Tamara, et al. "Streaming variational bayes." *Advances in Neural Information Processing Systems*. 2013.

-
- [Hoffman 13] Hoffman, Matthew D., et al. "Stochastic variational inference." *The Journal of Machine Learning Research* 14.1 (2013): 1303-1347.
- [Johnson 14] Johnson, Matthew, and Alan Willsky. "Stochastic variational inference for Bayesian time series models." *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*. 2014.
- [Taniguchi 16] Tadahiro Taniguchi, Ryo Nakashima, Hailong Liu and Shogo Nagasaka. "Double Articulation Analyzer with Deep Sparse Autoencoder for Unsupervised Word Discovery from Speech Signals." *Advanced Robotics* (2016) accepted.
- [Taniguchi 16] Tadahiro Taniguchi, Shogo Nagasaka, Ryo Nakashima. "Nonparametric Bayesian Double Articulation Analyzer for Direct Language Acquisition from Continuous Speech Signals." *IEEE Transactions on Cognitive and Developmental Systems* (2016) accepted.