

ガウス確率場による単語の意味変化と語義数の同時推定

井上 誠一¹ 小町 守¹ 小木曾 智信² 高村 大也³ 持橋 大地⁴

¹ 東京都立大学 ² 国立国語研究所 ³ 産業技術総合研究所 ⁴ 統計数理研究所
 inoue-seiichi@ed.tmu.ac.jp komachi@tmu.ac.jp togiso@ninjal.ac.jp
 takamura.hiroya@aist.go.jp daichi@ism.ac.jp

概要

本研究では、単語の通時的な意味変化のモデル化において、ディリクレ過程をガウス確率場の上に考えることで、単語によって異なる語義数を自動的に推定できる階層ベイズモデルを提案する¹⁾。疑似データを用いた実験において、意味変化と語義数を正しく推定できることを示した。また、実データを用いた実験においては、複数の解析対象の単語に対して推定結果を示し、分析した。

1 はじめに

言語は動的なシステムであり、常に進化し、話者とその環境の需要に適応している [1]。特に単語はさまざまな意味を持ち、その分布や広がりジャンルや文脈によって異なっている。例えば、“cute”という単語は18世紀初頭に登場し、もともと“賢い”という意味で使われていたが、19世紀後半に“狡猾な”という意味で使われ、現代においては、“魅力的な”という意味で使用されている [2]。このような通時的な単語の**意味変化**を捉えることができれば、辞書学での単語の意味変化に関する記述への利用や、単語の意味表現がより正確になることにより、意味情報が重要なタスクへの活用が可能になる。

近年では、単語の分散表現を用いて意味変化を検出する手法が数多く提案されているが [3, 4, 5]、これらは意味変化の検出はできるが、意味の趨勢や変化の様子を捉えることはできない。それに対し、解釈性の高い確率的生成モデルを用いて意味変化を捉える試みもある。Emms ら [6] は、単語の新たな意味の出現を捉えるための動的な生成モデルを提案しており、Frermann ら [7] は、意味の出現だけでなく変化のパターンも捉えることができるモデルを提案した。しかし、実際に単語の意味変化を調査した

い場合は解析対象の単語の**語義数**が自明であることは多くないにも関わらず、これらのモデルは語義数を事前に設定しなければならないという大きな問題がある。

そこで、我々はディリクレ過程をガウス確率場の上に考え Frermann らのモデルを拡張することで、語義数をデータから自動で推定し、単語の意味変化を捉えることのできる階層ベイズモデルを提案する。(1) 疑似データを用いた検証を行った結果、提案モデルが解析対象の単語の意味変化と語義数を正しく推定できることを定量的、定性的に示した。また、(2) 実データを用いた実験では、複数の単語に対し定性的な分析を行い、妥当性を示した。

2 関連研究

2.1 A Bayesian model of diachronic meaning change [7]

Frermann らは、単語の通時的な意味の発展を捉える、動的なベイズモデル (dynamic Bayesian model of Sense ChANge; 以下、SCAN と表記) を提案した。SCAN では、対象単語 w 一つに対して一つのモデルが構築され、入力、対象単語 w が含まれる文章の文脈単語集合 c で構成されるスニペットとその文章が出現した年のラベルとなっている。

SCAN において、時点 $t \in \{1 \dots T\}$ のスニペット集合は時点ごとのユニグラム混合:

- 意味上の K 次元多項分布 ϕ_t (意味分布)
- 各意味 k の語彙上の V 次元多項分布 $\psi_{t,k}$ (意味-単語分布)

でモデル化される。また、それぞれの事前分布にはガウス分布が仮定され、次のように変換することで ϕ を得る:

- 多次元ガウス分布から K 次元ベクトル α を生成
- ロジスティック変換 $\phi_k = \exp(\alpha_k) / \sum_{k=1}^K \exp(\alpha_k)$ によって $K-1$ 次元単体に射影。

1) 実装は <https://github.com/seiichiinoue/iscan> で公開している。

ψ についても同様である。そして、意味分布と意味-単語分布のパラメータ ϕ, ψ が、時間変化と共に変化をするように、事前分布に 1 階の内生的ガウス確率場 (iGMRF) [8] を定義する。iGMRF は、“近傍と似た値をとる” 事前分布であり²⁾、実数ベクトル $\mathbf{x} = (x_1, x_2, \dots, x_T)$ について、ガウス分布 $\mathcal{N}(\mu, \sigma^2)$ を用いて、次のように定義される:

$$x_t | \mathbf{x}_{-t}, \kappa \sim \mathcal{N}\left(\frac{1}{2}(x_{t-1} + x_{t+1}), \frac{1}{2\kappa}\right). \quad (1)$$

ただし、 \mathbf{x}_{-t} は \mathbf{x} から x_t を除いたものであり、 κ は精度パラメータである。また、意味分布と意味-単語分布の事前分布であるガウス分布は、それぞれ変化の度合いをコントロールするパラメータとして κ_ϕ, κ_ψ を持つ。特に、対象単語 w によって意味の変化の“速度”は異なるため、意味分布の事前分布の精度パラメータ κ_ϕ はデータから推定する。

これらを踏まえると、SCAN の生成モデルは次のようになる。Ga(a, b) はガンマ分布、Mult(θ) は多項分布を表す。

1. Draw $\kappa_\phi \sim \text{Ga}(a, b)$
2. For time interval $t = 1 \dots T$
 - (a) Draw sense distribution
 - i. $\alpha_t | \alpha_{-t}, \kappa_\phi \sim \mathcal{N}(\frac{1}{2}(\alpha_{t-1} + \alpha_{t+1}), \kappa_\phi^{-1})$
 - ii. $\phi_t = \text{Softmax}(\alpha_t)$
 - (b) For sense $k = 1 \dots K$
 - i. Draw sense-word distribution
 - A. $\beta_{t,k} | \beta_{-t}, \kappa_\psi \sim \mathcal{N}(\frac{1}{2}(\beta_{t-1,k} + \beta_{t+1,k}), \kappa_\psi^{-1})$
 - B. $\psi_{t,k} = \text{Softmax}(\beta_{t,k})$
 - (c) For snippet $d = 1 \dots D$
 - i. Draw sense $z_d \sim \text{Mult}(\phi_t)$
 - ii. For context position $i = 1 \dots I$
 - A. Draw word $w_{d,i} \sim \text{Mult}(\psi_{t,z_d})$

2.2 Logistic stick-breaking process [9]

ディリクレ過程は有限次元の多項分布を生成するディリクレ分布の無限次元版であり、その実現例の一つとして、stick-breaking process (SBP) [10] が挙げられる。SBP では、確率の総和である長さ 1 の棒を折っていくことで、 k 番目のクラスに対する確率 π_k を決定し、デルタ関数 $\delta(\theta_k)$ を基底測度 G_0 からサンプルした場所 $\theta_k \sim G_0$ に立てていくことで確率分布 G を生成しており、生成過程は次のようになる:

$$\pi_k = v_k \prod_{k'=1}^{k-1} (1 - v_{k'}), v_k \sim \text{Be}(1, a), \theta_k \sim G_0, \quad (2)$$

$$G = \sum_{k=1}^{\infty} \pi_k \delta(\theta_k). \quad (3)$$

2) カーネルが隣接する時間に限定されるガウス過程の特別な例と捉えることもできる。

ただし、Be($1, a$) はベータ分布である。

Ren ら [9] は、各クラスが何らかの共変量と紐づいている³⁾ 場合、それをロジスティック変換し各クラスの確率とすることで、同様にディリクレ過程を実現する logistic stick-breaking process (LSBP) を提案した。各クラスのガウス分布に従う確率変数を β_k としたとき、LSBP によって生成される確率分布 G_β は次のようになる:

$$\pi(\beta_k) = \sigma(\beta_k) \prod_{k'=1}^{k-1} (1 - \sigma(\beta_{k'})), \quad (4)$$

$$G_\beta = \sum_{k=1}^{\infty} \pi(\beta_k) \delta(\theta_k). \quad (5)$$

ただし、 $\sigma(x)$ はシグモイド関数を表す。

3 提案手法

3.1 Infinite SCAN

我々は、2.1 節で紹介した SCAN において、対象単語によって異なる語義数をコーパスから自動的に推定できるように、2.2 節で紹介した LSBP を用いて拡張を行った階層ベイズモデルを提案する。

提案モデルでは、SCAN と同様に対象単語 w 一つに対して一つのモデルを考え、入力も SCAN と同様に、コーパスにおける対象単語 w が含まれる文章の文脈単語集合 \mathbf{c} で構成されるスニペットとその文章が出現した年のラベルとする。

時点 t に出現したスニペット集合は、 ϕ_t と $\psi_{t,k}$ それぞれの上に定義される意味分布と意味-単語分布によって表され、それぞれがガウス分布に従う。ここで、意味分布のパラメータ ϕ_t について、対象単語 w によって異なる語義数 S_w をコーパスから自動的に推定できるように、LSBP を用いて生成過程を次のように変更する:

$$\alpha_t | \alpha_{-t}, \kappa_\phi \sim \mathcal{N}(\frac{1}{2}(\alpha_{t-1} + \alpha_{t+1}), \kappa_\phi^{-1}), \quad (6)$$

$$\phi_{t,k} = \sigma(\alpha_{t,k}) \prod_{k'=1}^{k-1} (1 - \sigma(\alpha_{t,k'})) \quad (k = 1, \dots, K). \quad (7)$$

ここで、LSBP は無限次元の多項分布を生成するが、実際には十分な語義の次元があればそれ以上使われることはないため、本研究では SCAN と同様に $K = 8$ とした⁴⁾。また、提案モデルでは、意味分布のパラメータ ϕ_t は、全体を見て正規化する Softmax 変換ではなく、それぞれの意味における停止確率を意味ごとに確率化する LSBP によって構築される。そ

3) SCAN の場合はガウス分布に従う確率変数が各意味、意味における各単語と紐づいている。

4) 事前実験より、8 個以上の語義を持つ単語はほとんど存在しなかった。

表 1 各モデルによって推定された意味分布と正解の意味分布の Kullback-Leibler 距離.

語義数 S_w	2	3	4	5
SCAN ($K = 8$)	0.2612	0.2788	0.1031	0.0085
Infinite SCAN	0.0009	0.0016	0.0044	0.0043

のため提案モデルでは, SCAN のように κ_ϕ を全ての意味 $k \in \{1 \dots K\}$ で共有するのではなく, 各意味の値 α ごとに異なる分散 $\kappa_\phi^{(k)}$ を仮定し推定する.

3.2 MCMC 法による推定

提案モデルの推定には, ブロック化 Gibbs sampler を用いる. 提案モデルにおける推定パラメータは, スニペットに割り当てられる意味 z , 意味分布と意味-単語分布の事前分布である α (ϕ の LSBP 変換前), β (ψ の Softmax 変換前), ガンマ分布に従う意味分布の事前分布の精度パラメータ κ_ϕ である. サンプルングの疑似コードを図 5 に示した.

スニペットの意味 各スニペットの意味 z_d は, サンプルング時点でのパラメータ ϕ, ψ を用いて, 次の事後分布に従ってサンプルングされる:

$$p(z_d | \mathbf{w}, t, \phi, \psi) \propto p(z_d | t) p(\mathbf{w} | t, z_d) = \phi_{z_d}^{(t)} \prod_{\mathbf{w} \in \mathbf{w}} \psi_{\mathbf{w}}^{(t, z_d)} \quad (8)$$

意味分布のパラメータ 意味分布の事前分布はガウス分布となっているため, 多項分布との間に共役性がないので, ディリクレ-多項分布のようなサンプルングはできない. Linderman ら [11] は, Pólya-gamma 補助変数を用いることで, LSBP を用いて表現される, ガウス分布を事前分布として持つ多項分布のパラメータに対するギブスサンプルングを提案しており, 本研究ではそれに従って推定を行う⁵⁾. α の事後分布は次のようになる:

$$p(\alpha_t | z, \alpha_{-t}, \omega) \propto \mathcal{N}(\omega^{-1} f(c) | \alpha_t) \mathcal{N}(\alpha_t | \alpha_{-t}, \kappa_\phi^{-1}) \\ \propto \mathcal{N}(\alpha_t | \tilde{\mu}, \tilde{\kappa}_\phi^{-1}). \quad (9)$$

ただし, ω は補助変数であり, c_k を k 番目の意味に属しているスニペットの個数, $N(c_k) = \sum_k c_k - \sum_{j < k} c_j$ として, Pólya-gamma 分布 $\omega | z, \alpha_t \sim \text{PG}(N(c_k), \alpha_t)$ からサンプルングされる. また, $f(c_k) = c_k - N(c_k)/2$ として, 事後分布の平均と分散は, $\tilde{\mu} = (f(c_k) + \mu_k \kappa_\phi) \cdot \tilde{\kappa}_\phi$, $\tilde{\kappa}_\phi = (\omega_k + \kappa_\phi)^{-1}$ である.

意味-単語分布のパラメータ 意味-単語分布も意味分布と同様に事前分布としてガウス分布を仮定しているため, ディリクレ-多項分布のようなサン

5) 詳細は参考文献 [11] を参照されたい.

表 2 実験に使用したコーパスの年代とサイズ.

コーパス	年代	単語数
COHA	1810–2009	142,587,656
CLMET3.1	1710–1810	4,531,505

リングはできない. Mimno ら [12] は, Softmax 変換を用いて表現される, ガウス分布を事前分布としても多項分布のパラメータに対するギブスサンプルングを提案しており, 本研究ではそれに従って推定を行う⁶⁾. スニペット数を D , スニペット長を N_d とすると, β の事後分布は次のようになる:

$$p(\beta_t) \propto \prod_{d=1}^D \left(\prod_{n=1}^{N_d} \frac{\exp(\beta_{w_n}^{(t, z_d)})}{1 + \sum_{w' \neq w} \exp(\beta_{w'}^{(t, z_d)})} \right) \mathcal{N}(\beta_t | \beta_{-t}, \kappa_\psi^{-1}) \quad (10)$$

精度パラメータ 平均が既知であるガウス分布の精度パラメータは, ガンマ事後分布に従う. ガンマ分布の形状パラメータを a , 尺度パラメータを b とすると, κ_ϕ の事後分布は次のようになる:

$$p(\kappa_\phi^{(k)}) = \text{Ga} \left(a + \frac{K}{2}, b + \frac{1}{2} \sum_{t=1}^T (\alpha_{t,k} - \bar{\alpha}_k) \right) \quad (11)$$

ただし, $\bar{\alpha}_k = \frac{1}{T} \sum_t \alpha_{t,k}$ は意味 k における α の平均である.

4 実験

4.1 疑似データによる検証

実データでの検証に先立ち, 提案モデルが, 任意の意味変化と語義数をきちんと推定できるか検証するため, 疑似データによる検証を行う. 疑似データの生成において, 意味分布の生成にはガウス過程を用い, 意味-単語分布の生成にはディリクレ分布を用いて正解データをランダムに生成した.

表 1 に, 語義数を $S_w = 2, \dots, 5$ と動かした場合における SCAN と提案モデルの疑似データに対する推定結果の比較を示した. ここでは, 意味の分布を正しく表現できるか定量的に測るため, 指標として推定された意味分布と正解の意味分布の $\mathbb{R}^{T \times K}$ 上での Kullback-Leibler 距離を用いた. 結果から, 語義数を自動的に推定できない SCAN と比較して提案モデルが優れていることがわかる. また, 図 1, 2 に, 語義数が $S_w = 5$ の場合の正解例と, 疑似データに対する提案モデルの推定結果を示した. 横軸が時点, 縦軸が時点における意味の確率を表し, 色に対応する意

6) 詳細は参考文献 [12] を参照されたい.

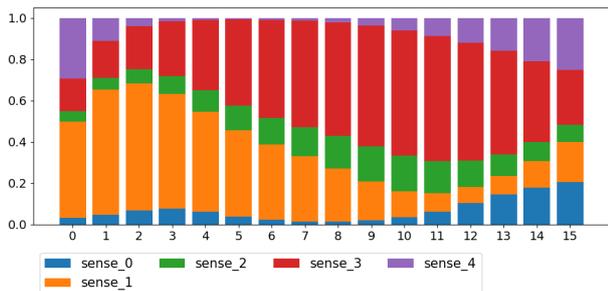


図1 語義数が $S_w = 5$ の場合の意味変化の正解例.

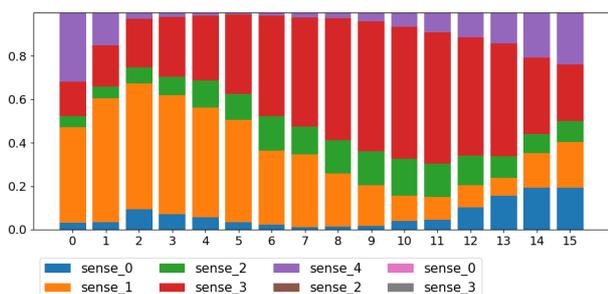


図2 語義数が $S_w = 5$ の疑似データに対する提案モデルの推定結果. 正解例に合わせて意味の並べ替えを行った.

味を凡例に示した. 図より, 提案モデルによる推定結果は意味の趨勢をほとんど正しく捉えていることがわかる.

4.2 データセット

本研究では, 英語を対象とし, 通時コーパスである Corpus of Historical American English (COHA) と The Corpus of Late Modern English Texts (CLMET3.1) を用いた. コーパスの統計量を表2に示した.

実験で用いる際は, トークナイズ, 見出し語化を行ったのち, ストップワードの削除を行った. また品詞タグ付けを行い, 名詞, 動詞, 形容詞のみを抽出し使用した. 以上の前処理を行ったのち, 対象単語が出現する箇所をコーパスから抽出し, 文脈窓幅を5として対象単語のコーパスを作成した. 対象単語としては, SCANで分析されたものを参考にして“coach”と“power”を選択した.

4.3 結果

図3, 4に“coach”と“power”に対する推定結果を示した. 凡例には, 意味 k における意味-単語分布 $\psi_k = \sum_i \psi_{i,k}$ の Normalized pointwise mutual information (NPMI) [13] の高い単語を上から10個示した.

“coach”では, コーパスから推定された語義数は2つとなった. 馬車や乗り物としての意味 (青色) か

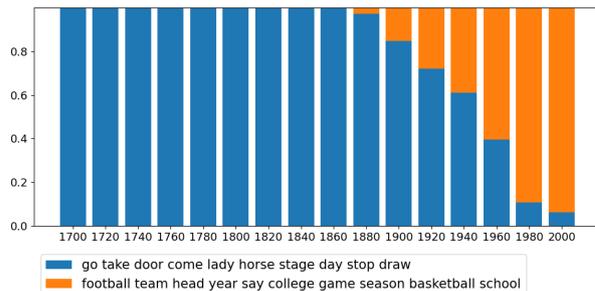


図3 対象単語“coach”に対する提案モデルの推定結果.

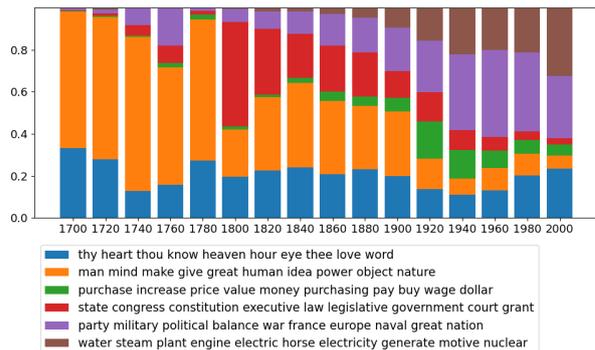


図4 対象単語“power”に対する提案モデルの推定結果.

ら指導や指導者という意味 (青色) に変化していることが捉えられている. “power”については, コーパスから推定された語義数は6つとなった. 宗教や教会の文脈で使用される力の意味 (青色) は, 現在に至っても使用されていることがわかる. また, 肉体上・精神上的の自然の能力, 体力, 知力といった意味 (青色) と法や政治的な力といった意味 (赤色) は現在に近づくにつれて使われることが少なくなっている. 一方で, 経済的な力としての意味 (緑色) や国際的な政治力や戦争などの文脈における力 (紫色), また動力の意味 (茶色) は現在になるにつれて使用されることが増えてきていることがわかる.

これらの結果は歴史的背景や辞書からも妥当であり, 提案モデルはコーパスからある程度正しく意味変化を捉えることができていることがわかる.

5 まとめと今後の展望

本研究では, 単語の通時的な意味変化のモデル化において, 単語によって異なる語義数を自動的に推定できる階層バイズモデルを提案した. 疑似データを用いた実験において, 意味変化と語義数を正しく推定できることを示した. また, 実データを用いた実験においては, 複数の解析対象の単語に対して推定結果を示し, 分析した. 今後は, 他の単語や英語以外の言語を対象とした分析を行っていききたい.

謝辞

本研究は国立国語研究所の共同研究プロジェクト「現代語の意味の変化に対する計算的・統計力学的アプローチ」, 「通時コーパスの設計と日本語史研究の新展開」および JSPS 科研費 19H00531, 18K11456 の研究成果の一部を報告したものである。

参考文献

- [1] Jean Aitchison. 2001. *Language Change: Progress Or Decay?*. Cambridge Approaches to Linguistics. Cambridge University Press.
- [2] Angus Stevenson, editor. 2010. *The Oxford English Dictionary*. Oxford University Press, third edition.
- [3] Robert Bamler, and Stephan Mandt. 2017. Dynamic Word Embeddings. In *Proceedings of the 34th International Conference on Machine Learning*, edited by Doina Precup and Yee Whye Teh, 70:380–89.
- [4] Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web, WWW'15*, p. 625–635.
- [5] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1489–1501.
- [6] Martin Emms, and Arun Kumar Jayapal. 2016. Dynamic Generative Model for Diachronic Sense Emergence Detection. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 1362–73.
- [7] Lea Frermann, and Mirella Lapata. 2016. A Bayesian Model of Diachronic Meaning Change. *Transactions of the Association for Computational Linguistics* 4: 31–45.
- [8] Havard Rue and Leonhard Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press.
- [9] Lu Ren, Lan Du, Lawrence Carin, and David Dunson. 2011. Logistic Stick-Breaking Process. *Journal of Machine Learning Research: JMLR* 12: 203–39.
- [10] Jayaram Sethuraman. 1994. A Constructive Definition of Dirichlet Priors. *Statistica Sinica* 4 (2): 639–50.
- [11] Scott Linderman, Matthew J. Johnson, and Ryan P. Adams. 2015. Dependent Multinomial Models Made Easy: Stick-Breaking with the Polya-Gamma Augmentation. In *Advances in Neural Information Processing Systems*, vol. 28, 3456–64.
- [12] David Mimno, Hanna Wallach, and Andrew McCallum. 2008. Gibbs Sampling for Logistic Normal Topic Models with Graph-Based Priors. In *Neural Information Processing Systems Workshop on Analyzing Graphs*. vol. 61.
- [13] Gerlof Bouma. Normalized (pointwise) mutual information in collocation extraction. 2009. *Proceedings of the German Society for Computational Linguistics and Lan-*

guage Technology, vol 30: 31–40.

- [14] Valerio Perrone, Marco Palma, Simon Hengchen, Alessandro Vatri, Jim Q. Smith, and Barbara McGillivray. 2019. GASC: Genre-Aware Semantic Change for Ancient Greek. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, 56–66.

A MCMC アルゴリズム

3.2 節で説明した提案モデルの推定に用いた Gibbs sampler の疑似コードを図 5 に示す. 各パラメータの初期値は基本的に Frermann ら [7] に従っている. ただし, 本研究では, 対象単語の意味の変化を“意味-単語分布の変化”ではなく, できるだけ“意味分布の変化”で捉えたいため (前者の場合, 語義数の推定が難しくなる), Perrone ら [14] に従い, κ_ψ の値を比較的大きな 100.0 とした.

Algorithm 1: MCMC Procedure

```

1 Initialize  $\kappa_\phi^{(k)} = 4.0$  (for all  $k$ )
2 Initialize  $\kappa_\psi = 100.0$ 
3 Initialize  $a = 7.0, b = 3.0$ 
4 for  $t = 1 \dots T$  do
5   Initialize  $\alpha_t \sim \mathcal{N}(\frac{1}{2}(\alpha_{t-1} + \alpha_{t+1}), \kappa_\phi^{-1})$ 
6   Set  $\phi_t = \text{LSB}(\alpha_t)$ 
7   for  $k = 1 \dots K$  do
8     Initialize  $\beta_{t,k} \sim \mathcal{N}(\frac{1}{2}(\beta_{t-1,k} + \beta_{t+1,k}), \kappa_\psi^{-1})$ 
9     Set  $\psi_{t,k} = \text{Softmax}(\beta_{t,k})$ 
10  end
11 end
12 for  $j = 1 \dots J$  do
13   Sample  $z$  according to Eq. (8)
14   for  $t = 1 \dots T$  do
15     Sample  $\phi$  according to posterior in Eq. (9)
16     Sample  $\psi$  according to posterior in Eq. (10)
17   end
18   Sample  $\kappa_\phi$  according to posterior in Eq. (11)
19 end

```

図 5 MCMC 法による Infinite SCAN の推定.

B 疑似データに対する推定結果

4.1 節で行った, 疑似データによる提案モデルの性能の検証において, 語義数 S_w が異なる場合の結果を示す. 図 6, 7 には語義数が $S_w = 2$ の正解例と推定結果, 図 8, 9 には語義数が $S_w = 3$ の正解例と推定結果を示した. 語義数が異なっても, 提案モデルは意味変化と語義数を適切に推定できていることがわかる.

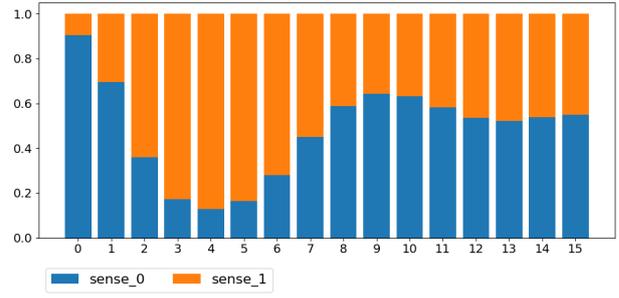


図 6 語義数が $S_w = 2$ の場合の正解例.

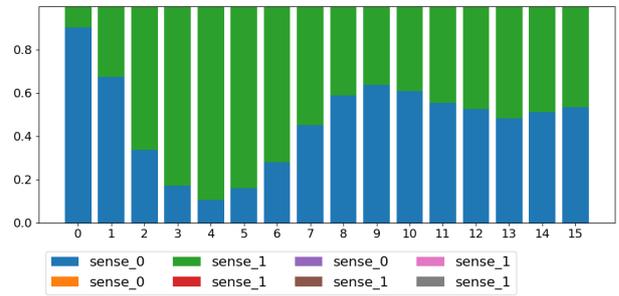


図 7 語義数が $S_w = 2$ の疑似データに対する提案モデルの推定結果. 正解例に合わせて意味の並べ替えを行った.

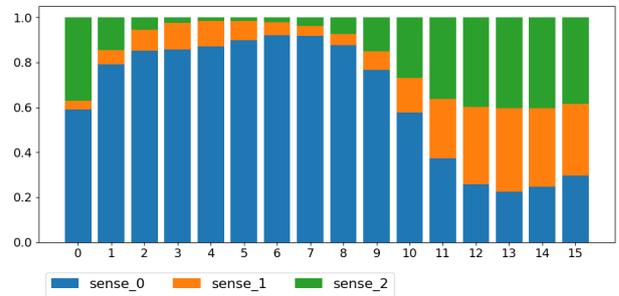


図 8 語義数が $S_w = 3$ の場合の正解例.

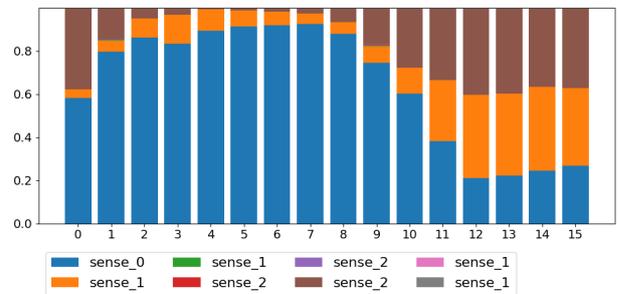


図 9 語義数が $S_w = 3$ の疑似データに対する提案モデルの推定結果. 正解例に合わせて意味の並べ替えを行った.