

# 木構造ニューラルトピックモデル

磯沼 大<sup>1</sup> 森 純一郎<sup>1</sup> ボレガラ ダヌシカ<sup>2</sup> 坂田 一郎<sup>1</sup><sup>1</sup> 東京大学 <sup>2</sup> リヴァプール大学

{isonuma, isakata}@ipr-ctr.t.u-tokyo.ac.jp mori@mi.u-tokyo.ac.jp

danushka@liverpool.ac.uk

## 1 はじめに

文書の潜在的なトピックを推定するトピックモデルは、Latent Dirichlet Allocation (LDA[2])をはじめ、情報検索や文書モデリングで頻用されている。その一種である木構造トピックモデル[8]は、無限の分岐を持つ木構造上において、根に一般的なトピックを、葉に近づくにつれ詳細なトピックを配置する(図1)。これにより多様な観点から簡明な情報を提供でき、観点付き感情分析[10]や抽出型要約[4]において有用である。

しかしながら、木構造トピックモデルはニューラルモデルを用いた下流タスクとの併用が困難である。ニューラルモデルの学習には大量の文書を要する一方、従来、事後分布推定に用いられている崩壊型ギブスサンプリング[8]や平均場近似[15]は、学習性能や並列化の困難さから、大規模な文書への適用が困難である。また、ニューラルモデルと、トピックモデルの事後分布を一体的に学習することで、下流タスクに有用なトピックとその木構造が得られるとより望ましい。

本研究では、文書からトピック分布への写像をニューラルネットワークにより構成した木構造ニューラルトピックモデルを提案する。提案モデルは autoencoding variational Bayes (AEVB[11]) に基づいた学習が可能であり、前述の問題を解決可能である。既存研究は LDA などのフラットなトピックモデルに AEVB を適用しているが[13, 14]、無限木上のトピック分布への写像の構築方法は自明でない。本研究は、親子間と兄弟間それぞれに再帰的な構造を持つ doubly-recurrent neural networks (DRNN[1]) を用いることで、文書から無限木上のトピック分布への写像を構築する。

実験から、提案モデルは、既存の木構造トピックモデル[3]とほぼ同等の一貫性を持つトピックと木構造を得ながら、より大規模な文書に適用できることが示された。また、ニューラルモデルを用いた観点付き感情分析[6]や要約生成[16]との一体的な学習が可能であり、木構造トピックモデルの応用可能性を広げた。

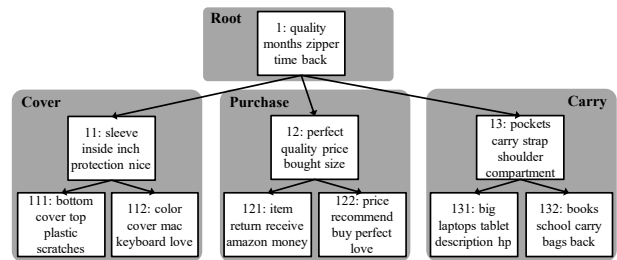


図 1: 提案モデルにより得られた Laptop Bags に関する商品レビューのトピックとその木構造。各トピックに最も頻出する 5 単語と人手で付与したラベルを表示。

## 2 関連研究

木構造トピックモデルの発案[8]以降、様々な拡張[15, 3, 9]と応用が提案されている。Goyal らは VAE の潜在表現の事前分布に木構造トピックモデルを用いて、ビデオのフレーム画像を木構造で表現した[7]。しかし、VAE の変数は勾配降下法で、トピックモデルの事後分布は平均場近似により個別に更新するため、少量のデータしか扱えず、学習には 1,241 件のビデオ(文書に相当)のみを用いている。本研究は、AEVB に基づく学習により、大量文書への適用と、下流タスクのニューラルモデルとの一体的な学習を可能にする。

AEVB に基づくフラットなトピックモデルとして、Miao らは recurrent neural networks (RNN) を用いて、動的な数のトピック分布を獲得している[13]。一方、本研究は DRNN により無限木上のトピック分布を得ることで、トピック間に潜在する木構造を獲得する。

## 3 木構造ニューラルトピックモデル

本節では、提案モデルによる文書の生成過程と AEVB に基づく学習方法について説明する。図 2 に示すように、提案モデルは各単語について、根ノードからいずれかの葉ノードまでのパスと、階層をサンプルする。そして、パス及び階層によって指定されたトピックに紐づく多項分布から、各単語をサンプルする。

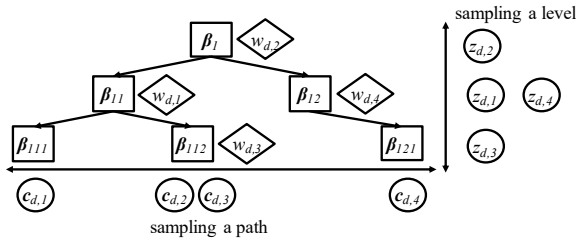


図 2: 各単語に紐づくトピックのサンプリング過程。

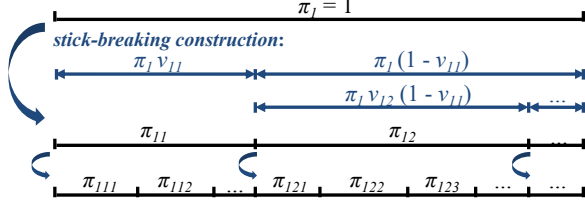


図 3: Tree-based stick-breaking construction の概要。

- 各文書インデックス  $d \in \{1, \dots, D\}$  について:
  - 正規確率ベクトルをサンプル:  $\mathbf{x}_d \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0^2)$  (1)
  - パス分布に変換:  $\boldsymbol{\pi}_d = f_\pi(\mathbf{x}_d)$  (2)
  - 階層分布に変換:  $\boldsymbol{\theta}_d = f_\theta(\mathbf{x}_d)$  (3)
- $d$  の各単語インデックス  $n \in \{1, \dots, N_d\}$  について:
  - パスをサンプル:  $c_{d,n} \sim \text{Mult}(\boldsymbol{\pi}_d)$  (4)
  - 階層をサンプル:  $z_{d,n} \sim \text{Mult}(\boldsymbol{\theta}_d)$  (5)
  - 単語をサンプル:  $w_{d,n} \sim \text{Mult}(\boldsymbol{\beta}_{c_{d,n}[z_{d,n}]})$  (6)

ただし、 $\boldsymbol{\beta}_{c_{d,n}[z_{d,n}]} \in \Delta^{V-1}$  はトピック  $c_{d,n}[z_{d,n}]$  に紐づく単語分布である。従来法 [15] は、文書毎にパスをサンプルし、パス上のトピックのみから文書を生成していた。一方、本研究は単語毎にパスをサンプルし、全てのトピックから文書を生成可能にする。

従来法 [15] は、パス及び階層分布を tree-based stick-breaking construction (図 3) に基づきサンプルする:

$$\nu_k \sim \text{Beta}(1, \gamma), \pi_k = \pi_{\text{par}(k)} \nu_k \prod_{j=1}^{k-1} (1 - \nu_j) \quad (7)$$

$$\eta_l \sim \text{Beta}(1, \alpha), \theta_l = \eta_l \prod_{j=1}^{l-1} (1 - \eta_j) \quad (8)$$

ただし、 $k \in \{1, \dots, K\}$  及び  $\text{par}(k)$  は  $k$  番目のトピックとその親を、 $l \in \{1, \dots, L\}$  は  $l$  番目の階層を示す。

一方、提案モデルは、正規確率ベクトルから無限木上のパス分布及び階層分布への写像  $f_\pi, f_\theta$  をニューラルネットワークにより構成し、AEVB に基づく事後分布推定を可能にする。特に無限木上のパス分布への写像を、DRNN を用いて構築する。

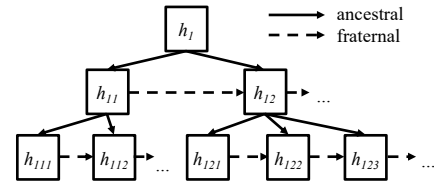


図 4: Doubly-recurrent neural networks の概要。

### 3.1 トピック分布の生成

図 4 に示すように、DRNN は親子間と兄弟間それぞれに再帰的な構造を持つ 2 つの RNN で構成される [1]。提案モデルは、この 2 つの再帰構造により、正規確率ベクトルから無限木上のパス分布への写像  $f_\pi$  を、有限の変数で表現する。トピック  $k$  の隠れ層  $\mathbf{h}_k$  と棒折り比率  $\nu_k$  は、DRNN を用いて下記式で計算される:

$$\mathbf{h}_k = \tanh(\mathbf{W}_p \mathbf{h}_{\text{par}(k)} + \mathbf{W}_s \mathbf{h}_{k-1}) \quad (9)$$

$$\nu_k = \text{sigmoid}(\mathbf{h}_k^\top \mathbf{x}) \quad (10)$$

ここで、 $\text{par}(k)$  及び  $k-1$  は、それぞれトピック  $k$  の親と一つ手前の兄弟を示す。正規確率ベクトルから階層分布への写像  $f_\theta$  も、同様に RNN で構成する:

$$\mathbf{h}_l = \tanh(\mathbf{W}_l \mathbf{h}_{l-1}) \quad (11)$$

$$\eta_l = \text{sigmoid}(\mathbf{h}_l^\top \mathbf{x}) \quad (12)$$

### 3.2 単語分布の生成

トピック  $k$  及び単語の潜在表現  $\mathbf{t}_k \in \mathbb{R}^H$ ,  $\mathbf{U} \in \mathbb{R}^{V \times H}$  を導入し、単語分布  $\boldsymbol{\beta}_k \in \Delta^{V-1}$  を得る。本研究では動的な数のトピックを扱うため、もう一つの DRNN により、トピックの潜在表現も動的に生成する:

$$\mathbf{t}_k = \text{DRNN}(\mathbf{t}_{\text{par}(k)}, \mathbf{t}_{k-1}) \quad (13)$$

$$\boldsymbol{\beta}_k = \text{softmax}\left(\frac{\mathbf{U} \cdot \mathbf{t}_k^\top}{\tau^{\frac{1}{l}}}\right) \quad (14)$$

ここで、 $\tau^{\frac{1}{l}}$  は温度パラメータであり、トピックの深さ  $l$  が浅いほど、よりなだらかな単語分布を生成する。

フラットなニューラルトピックモデルでは、トピック間の冗長性を排除するために、全トピックの単語分布を独立にする正則化項が導入される [13]。一方で、木構造トピックモデルは親子間の単語分布の相関を許容するため、本研究では下記の正則化項を導入する:

$$\sum_{k \notin \text{Leaf}} \sum_{i, j \in \text{Chi}(k): i \neq j} \left( \frac{\bar{\mathbf{t}}_{ki}^\top \cdot \bar{\mathbf{t}}_{kj}}{\|\bar{\mathbf{t}}_{ki}\| \|\bar{\mathbf{t}}_{kj}\|} - 1 \right)^2 \quad (15)$$

ただし、 $\bar{\mathbf{t}}_{ki} = \mathbf{t}_i - \mathbf{t}_k$  は親と子の単語分布の差分を、Leaf 及び  $\text{Chi}(k)$  は最下層のトピックと、トピック  $k$  の子の集合をそれぞれ表す。これにより、トピックの冗長性を排しつつ、単語分布の親子間の相関を許容する。

### 3.3 AEVBに基づく事後分布の推定

提案モデルによる文書の尤度は下記式で表される。

$$\begin{aligned}
 p(\mathbf{w}_{1:N_d} | \boldsymbol{\mu}_0, \boldsymbol{\sigma}_0, \boldsymbol{\beta}) &= \int_{\boldsymbol{\pi}, \boldsymbol{\theta}} \left\{ \prod_n \sum_{\mathbf{c}_n, z_n} p(w_n | \boldsymbol{\beta}_{\mathbf{c}_n[z_n]}) p(\mathbf{c}_n | \boldsymbol{\pi}) p(z_n | \boldsymbol{\theta}) \right\} \\
 &\quad p(\boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{\mu}_0, \boldsymbol{\sigma}_0) d\boldsymbol{\pi} d\boldsymbol{\theta} \\
 &= \int_{\boldsymbol{\pi}, \boldsymbol{\theta}} \left\{ \prod_n (\boldsymbol{\beta} \cdot \boldsymbol{\phi})_{w_n} \right\} p(\boldsymbol{\pi}, \boldsymbol{\theta} | \boldsymbol{\mu}_0, \boldsymbol{\sigma}_0) d\boldsymbol{\pi} d\boldsymbol{\theta}
 \end{aligned} \tag{16}$$

ここで  $\boldsymbol{\phi} \in \Delta^{K-1}$  は文書のトピック分布で、 $\phi_k = \sum_{l=1}^L \theta_l (\sum_{c:c_l=k} \pi_c)$  により計算する。本研究では、AEVBに基づき文書のBOW表現  $\mathbf{w}_d$  をガウス分布に変換する多層パーセプトロン (MLP)  $f_\mu$  と  $f_\sigma$  を導入し、トピックの変分分布を構築する。変分分布から  $\hat{\boldsymbol{\phi}}$  をサンプルして式 (17) に示す文書の対数尤度の変分下限を計算し、それを最大化する変分分布を得る。

$$\begin{aligned}
 \mathcal{L}_d \approx \sum_n \log(\boldsymbol{\beta} \cdot \hat{\boldsymbol{\phi}})_{w_n} - \\
 \text{KL}[\mathcal{N}(\mathbf{x} | f_\mu(\mathbf{w}_d), f_\sigma^2(\mathbf{w}_d)) || \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_0, \boldsymbol{\sigma}_0^2)]
 \end{aligned} \tag{17}$$

### 3.4 動的な木構造の更新

動的な木構造を可能にするため、本研究は2つのヒューリスティクスを導入し、トピックを追加・消去する。まず、トピック  $k$  に含まれる単語数の割合を  $p_k = (\sum_{d=1}^D N_d \hat{\phi}_{d,k}) / (\sum_{d=1}^D N_d)$  で計算する。葉でない全てのトピック  $k$  について、 $p_k$  が閾値より大きい場合、より明瞭なトピックを得るために子を追加する。全てのトピック  $k$  について、子孫全体に含まれる単語数の割合  $\sum_{j \in \text{Des}(k)} p_k$  が閾値よりも小さい場合、トピック  $k$  とその子孫を消去する (Des( $k$ ) はトピック  $k$  の子孫の集合を示す)。また、トピックが最下層に子を持たなくなった場合、当該トピックも削除する。

## 4 実験

### 4.1 データセット

本実験では、20NewsGroups<sup>1</sup> 及び Amazon product reviews<sup>2</sup> をデータセットとして用いた。20NewsGroups は 20 ドメインのニュースから構成され、11,258 件の学習用文書と、7,487 件の評価用文書からなる。Amazon product reviews については、Laptop Bags のレビューを利用し、学習、検証及び評価用文書の件数は、それぞれ 31,943 件、385 件及び 416 件である。

<sup>1</sup>[https://github.com/akashgit/autoencoding\\_vi\\_for\\_topic\\_models](https://github.com/akashgit/autoencoding_vi_for_topic_models)

<sup>2</sup><https://github.com/stangelid/oposum>

Models	20News	Amazon
RSB [13]	0.201	0.102
nCRP [3]	0.198	0.112
Our Model	0.220	0.121

表 1: 各モデルにより得られたトピックの平均 NPMI。

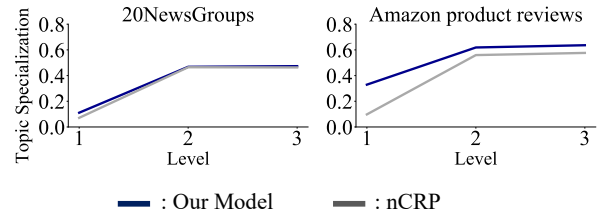


図 5: 各階層の平均トピック詳細度。

### 4.2 ベースライン

ベースラインとして、nested Chinese restaurant process に基づく木構造トピックモデル (nCRP[3]) を用いた。また、AEVBに基づきフラットなニューラルトピックモデルとして、文書から動的な数のトピック分布への写像を RNN で構成する Recurrent Stick Breaking process (RSB[13]) を採用した。

### 4.3 実装詳細

各モデルのハイパーパラメータは、Amazon product reviews の検証用文書を用いて以下のように設定した。枝の追加・削除の閾値 (3.4 節) は共に 0.05 であり、温度パラメータ (3.2 節) は  $\tau=10$  である。木の深さは 3 に設定し、第 2 層及び第 3 層のトピック数の初期値はいずれも 3 である。単語の分散表現、MLP 及び RNN の隠れ層の次元数は 256 で、勾配降下法は AdaGrad (学習率:  $10^{-2}$ 、バッチ毎の文書件数: 64) を利用した。nCRP の集中度は  $\gamma = 0.01$ 、ディリクレ分布の集中度は  $\eta = 5$ 、GEM 分布のハイパーパラメータは  $\pi = 10, m = 0.5$  である。

### 4.4 トピックの解釈可能性に関する評価

パープレキシティはトピックの解釈可能性の評価に不適切なことが指摘されている一方 [5]、Lau らはトピック内の各単語対に関する正規化自己相互情報量 (NPMI) が、人手による評価と強く相関することを示している [12]。本実験も既存研究 [13, 14] と同様に、NPMI によりトピックの解釈可能性を評価した。各モデルについて、全トピックの平均 NPMI を表 1 に示す。提案モデルは nCRP 及び RSB 対しいずれのデータセットにおいても競合し、ベースラインと同程度に解釈可能なトピックを得られることが確認された。

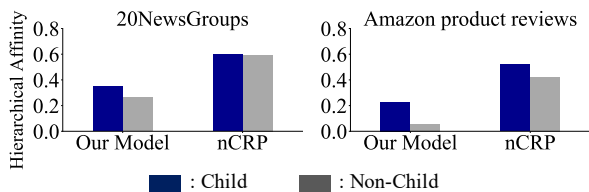


図 6: 親子と親子でないトピック間の類似度。

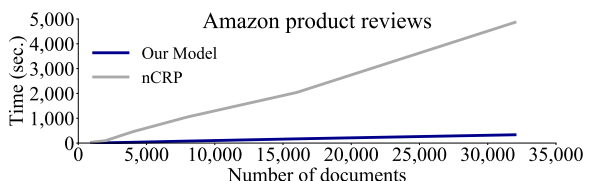


図 7: 学習用文書の件数に対する各モデルの学習時間。

#### 4.5 木構造の性質に関する評価

以下に示す 2 つの木構造の評価指標 [9] をもとに、木構造の性質を定量評価した。また、得られたトピックと木構造の例を 1 節の図 1 に示す。

**トピック詳細度:** 木構造の重要な特徴として、最も抽象的なトピックが根に配置され、葉に近づくにつれトピックが詳細になる性質が挙げられる。そこでトピック詳細度を、コーパス全体の単語分布とトピックの単語分布との類似度として定量化する。コーパス全体は最も一般的なトピックとしてみなせるため、より詳細なトピックは低い類似度を示す。各階層におけるトピック詳細度の平均を図 5 に示す。全体傾向は両者ともに類似することが確認された。

**親子間類似度:** もう一つの望ましい性質として、親のトピックが子のトピックに類似することが挙げられる。当該性質を定量化するため、第 2 層の各トピックについて、子トピックと子でないトピックとの類似度をそれぞれ計算し、その平均を図 6 に示す。提案モデルは nCRP より、親トピックにより類似する子トピックを生成していることが確認できる。また、提案モデルの低いスコアは、nCRP よりもより多様なトピックを得られることを示唆している。

#### 4.6 学習時間に関する評価

提案モデルが大量文書に適用できることを検証するため、Amazon product reviews の学習用文書の件数を変化させ、それぞれの学習時間を計測した<sup>3</sup>。図 7 に示すように、提案モデルは文書の件数に対する学習時間の増加は nCRP より緩やかで、32,000 件の文書では学習が 15 倍速いことが確認された。

<sup>3</sup>検証用文書のパープレキシティが 10 バッチ連続で改善しなかった時点で学習を終了した。CPU は Xeon E5-2683-v4 (2.1 GHz, 16 cores) を、GPU は GeForce GTX 1080 (8GB) を使用。

## 5 おわりに

本研究は、無限木構造上のトピック分布を doubly-recurrent neural networks により構築する木構造ニューラルトピックモデルを提案した。評価実験の結果、提案モデルは、従来法 [3] とほぼ同等の一貫性を持つトピックと木構造を得ながら、より大規模な文書に適用できることが示された。提案モデルは、ニューラルモデルを用いた下流タスクとの一体的な学習が可能であり、より広範なタスクへの応用が期待される。

## 参考文献

- [1] D. Alvarez-Melis and T. S. Jaakkola. Tree-structured decoding with doubly-recurrent neural networks. *ICLR*, 2017.
- [2] D. M. Blei, et al. Latent dirichlet allocation. *JMLR*, Vol. 3, pp. 993–1022, 2003.
- [3] D. M. Blei, et al. The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies. *JACM*, Vol. 57, p. 7, 2010.
- [4] A. Celikyilmaz and D. Hakkani-Tur. A hybrid hierarchical model for multi-document summarization. *ACL*, pp. 815–824, 2010.
- [5] J. Chang, et al. Reading tea leaves: How humans interpret topic models. *NIPS*, pp. 288–296, 2009.
- [6] B. Esmaeili, et al. Structured neural topic models for reviews. *AISTATS*, pp. 3429–3439, 2019.
- [7] P. Goyal, et al. Nonparametric variational autoencoders for hierarchical representation learning. *ICCV*, pp. 5094–5102, 2017.
- [8] T. L. Griffiths, et al. Hierarchical topic models and the nested chinese restaurant process. *NIPS*, pp. 17–24, 2004.
- [9] J. H. Kim, et al. Modeling topic hierarchies with the recursive chinese restaurant process. *CIKM*, pp. 783–792, 2012.
- [10] S. Kim, et al. A hierarchical aspect-sentiment model for online reviews. *AAAI*, pp. 526–533, 2013.
- [11] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *ICLR*, 2014.
- [12] J. H. Lau, et al. Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. *EACL*, pp. 530–539, 2014.
- [13] Y. Miao, et al. Discovering discrete latent topics with neural variational inference. *ICML*, pp. 2410–2419, 2017.
- [14] A. Srivastava and C. Sutton. Autoencoding variational inference for topic models. *ICLR*, 2017.
- [15] C. Wang and D. M. Blei. Variational inference for the nested chinese restaurant process. *NIPS*, pp. 1990–1998, 2009.
- [16] W. Wang, et al. Topic-guided variational autoencoder for text generation. *NAACL-HLT*, Vol. 1, pp. 166–177, 2019.