Vol.2020-NL-246 No.31
Vol.2020-SLP-134 No.31
2020/12/3

潜在的なトピック構造を捉えた生成型教師なし意見要約

磯沼 大[1,a)]　森 純一郎[1]　ボレガラ ダヌシカ[2]　坂田 一郎[1]

概要：本研究は商品レビューなどの意見文書を対象にした生成型教師なし要約手法を提案する。生成型教師なし要約では、参照要約なしに要約文の潜在表現をいかに獲得するかが鍵となる。そこで本研究では、根は一般的なトピックを、葉に近づくにつれより詳細なトピックを持つトピック木を導入し、意見文書の要約の各文が木構造上のトピックに対応することに着目した。文書から木構造上のトピックを推定し、トピック毎に要約文を生成することで、意見文書の要約が教師なしに得られることを示す。要約生成評価実験において、提案法は最新の教師なし要約生成手法と競合する性能を持つことを示した。また、根の文の潜在分布は分散が大きく一般的な文が生成される一方、葉に近づくにつれ分散が小さくなり具体的な文が生成されるといった特性を確認した。これは「動物」といったタクソノミー上の上位語の潜在分布は分散が大きく、「犬」や「猫」といった下位語は分散が小さくなるという、Gaussian word embedding に類似する特性であり、質問応答や対話生成などの文の詳細度合いを考慮する他タスクにも有用な知見である。

Unsupervised Opinion Summary Generation by Inducing Latent Topic Structure

1. はじめに

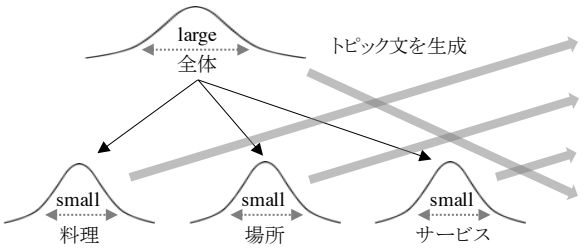
近年、EC サイト上の商品レビューや SNS 上の投稿の急激な増加により、それらの意見を機械で要約し、ニーズや世論を俯瞰する取り組みが注目されている。従来自動要約では抽出型アプローチが広く用いられているが，特に意見文書では必要十分な内容を捉えるのが困難であることが報告されている [5]。一方、生成型要約は言い換えや一般化を交えることで過不足のない要約を生成できることから、より有効なアプローチとして期待されている [14]。生成型要約のうち、教師ありアプローチは近年飛躍的な性能向上を遂げているものの、それらの適用先は大量の参照要約が利用可能なニュース記事など特定のドメインに限定されている。一方、意見文書はドメインが多様であり、大量の参照要約を人手で用意するには困難なことから、近年意見文書に対する教師なしアプローチが着目されている。

教師なしアプローチでは、要約の潜在表現を参照要約なしにいかに獲得するかが鍵となるが、本研究ではトピックとその構造を捉えることで要約文を生成する。例えば図 1 に示したあるレビューの要約は、「料理」、「サービス」、「場所」の各トピックについて詳述し、最後に全体的な印象を述べている。このように、要約は多様なトピックで構成されており、あるトピックは詳述され、他のトピックは簡潔に記述されていることが観察される。そこで本研究では、文書から木構造上のトピックを推定し、根からは全体的な、葉に近づくにつれより詳細なトピックに関する要約文（**トピック文**）を生成する。それらトピック文から、要約として相応しいトピックと詳細度合いを持つ文を選択することで、意見文書の要約が得られることを示す。

トピック文生成の文脈では、Wang らは文書中の文の潜在分布を混合ガウス分布 (**GMM**) で表現すると、その構成要素である各単峰ガウス分布がトピック文の潜在分布に相当することを明らかにした [26]。一方、本研究のように多様な詳細度合いを持つトピック文を生成するためには、文の詳細度合いを潜在空間上でモデル化する必要があり、その方法は明らかでない。そこで本研究では、トピック文の潜在表現をガウス分布で表現する際に、子の分布の分散が親よりも小さくなるようにモデルを構築することで、葉に近づくにつれより詳細な文を生成する（図 1）。単語の潜在分布としてガウス分布を用いる Gaussian word embedding では、「犬」のような具体的な単語は、「動物」といった一般的な単語よりも小さい分散を持つことが示されている [24]。

[1] 東京大学
[2] リヴァプール大学
[a)] isonuma@ipr-ctr.t.u-tokyo.ac.jp

Vol.2020-NL-246 No.31
Vol.2020-SLP-134 No.31
2020/12/3



要約（トピック文）

The food here is fantastic, easily the best sub sandwiches in the Arizona area.

The shop is local and family run, so I definitely choose it over a lot of the large national chains that are all around town.

The staff are extremely friendly and will always go above and beyond in creating a delicious sandwich for you.

You will not be let down by the great food that they make here!

図 1 本研究の概要図。右の要約はあるレストランのレビューについて実際に人が作成した要約であり、各文が木構造上の各トピックに対応することがわかる。トピック文の潜在分布の分散は葉に近づくほど小さくなり、より具体的なトピックに関する文に対応する。

文においても同様に、具体的な文は意味の分散が小さいため、その潜在分布は分散が小さいと考えられる。子の分布の分散を親よりも小さくするために、本研究では再帰的混合ガウス分布（**再帰的 GMM**）を文書中の文の潜在表現の事前分布として導入する。再帰的 GMM は、木構造上の各トピックに対応するガウス分布で構成され、子の事前分布に親の事後分布を設定することで構築される。これにより、根の潜在分布の分散は大きく、葉に近づくにつれ分散は小さくなる。実験では、既存の教師なし生成型要約手法と競合する要約性能を確認し、得られたトピック文が要約として機能することを示した。また、トピック文の詳細度合いはその潜在分布の分散の大きさに依存し、分散が小さいほどより具体的なトピック文が生成されることを示した。

2. 関連研究

2.1 教師なし生成型要約

従来、教師なし要約手法は抽出型アプローチに主眼を置いていたが [12], [13]、特に意見文書では抽出型要約ではその内容を網羅的に捉えることが困難なことから、意見文書に対する教師なし生成型要約が近年取組まれ始めている。複数文書の教師なし生成型要約の先駆的な手法である MeanSum は各文書の潜在表現の平均を算出し、それを自然言語に復号することで教師なし要約生成を実現した [9]。これを拡張した手法である Copycat は、製品やサービス毎に潜在表現の平均を得ることで、各製品・サービスに特有の内容を要約に含めることを可能にした。Ammplayo らは、文書にノイズを加えて擬似文書を作成し、擬似文書を要約対象文書、元の文書を要約とみなして学習することで、教師なし要約が可能になることを示した [1]。これら一連のアプローチは、いずれもトピックやその詳細度合いを考慮せずに、共通意見を網羅的に捉えている。一方、提案法は文書に潜在するトピックとその詳細度合いを明示的に捉えることで、重要なトピックに焦点を当てた要約を生成する。抽出型要約においては、トピック [2], [23] やその木構造 [6], [7] の有効性が従来研究で示されているが、本研究はそれらの教師なし生成型要約への応用を提案する。

2.2 文の生成モデルとトピック文生成

Bowman らは変分自己符号化器（**VAE**）を用いて文の潜在表現をベクトルではなくガウス分布で獲得している [3]。これにより、2 つの文の潜在表現の中間サンプルから、文法的な誤りが少なく、かつ 2 つの文と関連するトピックを持つ文を復号できることを明らかにした。具体的には文書の生成過程を以下のように仮定している。

各文書インデックス $d \in \{1, \ldots, D\}$ について:

文書 d 中の各文インデックス $s \in \{1, \ldots, S_d\}$ について:

1. 文の潜在表現 $\boldsymbol{x}_s \in \mathcal{R}^H$ をサンプル:

$$\boldsymbol{x}_s | z_s \sim p(\boldsymbol{x}_s) \tag{1}$$

2. 尤度が最大となる文 $\boldsymbol{w}_s$ を得る。

$$\boldsymbol{w}_s | \boldsymbol{x}_s = \arg\max_{\boldsymbol{w}_s} p(\boldsymbol{w}_s | \boldsymbol{x}_s) \tag{2}$$

ただし、$p(\boldsymbol{w}_s|\boldsymbol{x}_s) = \prod_t p(w_s^t | \boldsymbol{w}_s^{<t}, \boldsymbol{x}_s)$ は再帰的ニューラルネットワーク（**RNN**）デコーダにより求められ、文の潜在表現の事前分布は標準ガウス分布である: $p(\boldsymbol{x}_s) = \mathcal{N}(\boldsymbol{x}_s | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$。文書の尤度とその対数の変分下限は下記式で表される。

$$p(\boldsymbol{W}_{1:S_d}) = \prod_{s=1}^{S_d} \Big\{ \int p(\boldsymbol{w}_s|\boldsymbol{x}_s) p(\boldsymbol{x}_s) d\boldsymbol{x}_s \Big\} \tag{3}$$

$$\mathcal{L}_d = \sum_{s=1}^{S_d} \Big\{ \mathbf{E}_{q(\boldsymbol{x}_s|\boldsymbol{w}_s)} \left[\log p(\boldsymbol{w}_s|\boldsymbol{x}_s) \right] - \mathrm{D}_{\mathrm{KL}} \left[q(\boldsymbol{x}_s|\boldsymbol{w}_s) | p(\boldsymbol{x}_s) \right] \Big\} \tag{4}$$

$$q(\boldsymbol{x}_s|\boldsymbol{w}_s) = \mathcal{N}(\boldsymbol{x}_s | \hat{\boldsymbol{\mu}}_s, \hat{\boldsymbol{\Sigma}}_s), \hat{\boldsymbol{\mu}}_s = f_\mu(\boldsymbol{w}_s), \hat{\boldsymbol{\Sigma}}_s = \mathrm{diag}(f_\Sigma(\boldsymbol{w}_s)) \tag{5}$$

ただし $\hat{\boldsymbol{\Sigma}}_s$ は対角行列であり、f_μ と f_Σ は RNN エンコーダである。

これを拡張した取り組みとして、Wang らは文の事前分布としてガウス分布ではなく GMM を用いた: $p(\boldsymbol{x}_s) = \sum_k p(z_s = k) p(\boldsymbol{x}_s | z_s = k) = \sum_k p(z_s = k) \mathcal{N}(\boldsymbol{x}_s | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$。ただし、$\mathcal{N}(\boldsymbol{x}_s | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ はトピック k に対応するトピック文の潜在分布であり、この潜在分布からのサンプルを復号することでトピック文が生成できることを明らかにした。

一方、本研究は事前分布を再帰的 GMM で構築し、木構造上の各トピックに対応するトピック文を生成する。

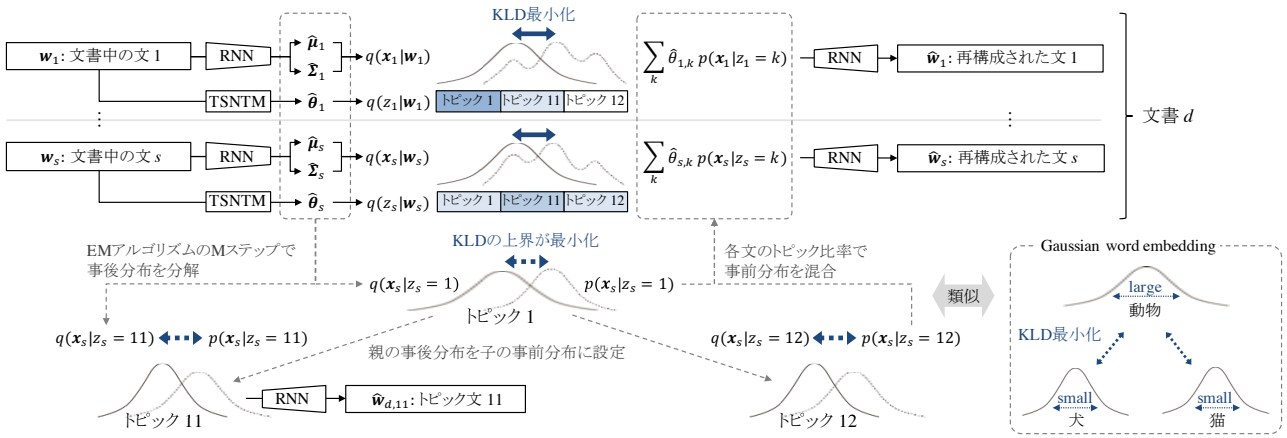


図 2 提案法の概要図。文書の文の事前分布に再帰的 GMM が設定され、その構成要素である単峰ガウス分布が各トピック文の事前分布に対応する。

3. 提案法

本節では、文書からトピック文を生成し、要約を得る過程について説明する。提案法の概要を図 2 に示す。

3.1 文書の生成過程

提案法では以下のように文書の生成過程を仮定する。

各文書インデックス $d \in \{1, \dots, D\}$ について:

文書 d 中の各文インデックス $s \in \{1, \dots, S_d\}$ について:

1. 文のトピック $z_s \in \{1, \dots, K\}$ をサンプル:
$$z_s \sim \mathrm{Mult}(\boldsymbol{\theta}) \tag{6}$$
2. 文の潜在表現 $\boldsymbol{x}_s \in \mathcal{R}^H$ をサンプル:
$$\boldsymbol{x}_s | z_s \sim \prod_{k=1}^{K} p(\boldsymbol{x}_s | z_s = k)^{\delta(z_s = k)} \tag{7}$$
3. 尤度が最大となる文 $\boldsymbol{w}_s$ を得る。
$$\boldsymbol{w}_s | \boldsymbol{x}_s = \arg\max_{\boldsymbol{w}_s} p(\boldsymbol{w}_s | \boldsymbol{x}_s) \tag{8}$$

ただし、$p(\boldsymbol{w}_s | \boldsymbol{x}_s)$ は RNN デコーダにより得られる。トピック分布は木構造上に定義され、既存研究 [11], [27] と同様に、一様分布をその事前分布に設定する：$\theta_k = K^{-1}$。文書中の文の事前分布として再帰的 GMM を設定し (7)、各要素はトピック文の事前分布 $p(\boldsymbol{x}_s | z_s = k)$ に対応する：

$$p(\boldsymbol{x}_s | z_s = 1) = \mathcal{N}(\boldsymbol{x}_s | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \tag{9}$$

$$\begin{aligned} p(\boldsymbol{x}_s | z_s = k) &= q(\boldsymbol{x}_s | z_s = par(k)) \\ &= \mathcal{N}(\boldsymbol{x}_s | \hat{\boldsymbol{\mu}}_{d,par(k)}, \hat{\boldsymbol{\Sigma}}_{d,par(k)}) \ (k \neq 1) \end{aligned} \tag{10}$$

ただし、$par(k)$ はトピック k の親を指す。$q(\boldsymbol{x}_s | z_s = par(k))$ は親トピック文の推定事後分布であり、子トピック文の事前分布として設定される。

文書の尤度とその対数の変分下限は下記式で表される。

$$p(\boldsymbol{W}_{1:S_d}) = \prod_{s=1}^{S_d} \left\{ \int p(\boldsymbol{w}_s | \boldsymbol{x}_s) p(\boldsymbol{x}_s | z_s) p(z_s) d\boldsymbol{x}_s dz_s \right\} \tag{11}$$

$$\begin{aligned} \mathcal{L}_d = \sum_{s=1}^{S_d} \Big\{ & \mathbf{E}_{q(\boldsymbol{x}_s|\boldsymbol{w}_s)} \left[\log p(\boldsymbol{w}_s | \boldsymbol{x}_s) \right] \\ & - \mathbf{E}_{q(\boldsymbol{x}_s|\boldsymbol{w}_s) q(z_s|\boldsymbol{w}_s)} \left[\log q(\boldsymbol{x}_s | \boldsymbol{w}_s) - \log p(\boldsymbol{x}_s | z_s) \right] \\ & - \mathbf{E}_{q(z_s|\boldsymbol{w}_s)} \left[\log q(z_s | \boldsymbol{w}_s) - \log p(z_s) \right] \Big\} \\ = \sum_{s=1}^{S_d} \Big\{ & \mathbf{E}_{q(\boldsymbol{x}_s|\boldsymbol{w}_s)} \left[\log p(\boldsymbol{w}_s | \boldsymbol{x}_s) \right] - \mathrm{D}_{\mathrm{KL}} \left[q(z_s | \boldsymbol{w}_s) | p(z_s) \right] \Big\} \\ & - \sum_{k=1}^{K} \sum_{s=1}^{S_d} \Big\{ \hat{\theta}_{s,k} \mathrm{D}_{\mathrm{KL}} \left[q(\boldsymbol{x}_s | \boldsymbol{w}_s) | p(\boldsymbol{x}_s | z_s = k) \right] \Big\} \end{aligned} \tag{12}$$

ただし $q(\boldsymbol{x}_s | \boldsymbol{w}_s) = \mathcal{N}(\boldsymbol{x}_s | \hat{\boldsymbol{\mu}}_s, \hat{\boldsymbol{\Sigma}}_s)$ は文 s の潜在表現の事後分布で、式 (5) と同様に RNN エンコーダにより推定する。$\hat{\theta}_{s,k} = q(z_s = k | \boldsymbol{w}_s)$ はトピック分布の事後分布であり、本研究では木構造ニューラルトピックモデル（**TSNTM** [15]）により推定する（6.1 節にて詳述）。

3.2 トピック文の生成

トピック文の潜在表現の事後分布は式 (13) で与えられ、$\sum_{s=1}^{S_d} \mathbf{E}_{q(\boldsymbol{x}_s|\boldsymbol{w}_s) q(z_s|\boldsymbol{w}_s)} \left[\log q(\boldsymbol{x}_s | z_s) \right]$ を最大化するパラメータを EM アルゴリズムの M ステップで導出する。

$$q(\boldsymbol{x}_s | z_s) = \prod_{k=1}^{K} \mathcal{N}(\boldsymbol{x}_s | \hat{\boldsymbol{\mu}}_{d,k}, \hat{\boldsymbol{\Sigma}}_{d,k})^{\delta(z_s = k)} \tag{13}$$

$$\hat{\boldsymbol{\mu}}_{d,k} = \frac{\sum_{s=1}^{S_d} \hat{\theta}_{s,k} \hat{\boldsymbol{\mu}}_s}{\sum_{s=1}^{S_d} \hat{\theta}_{s,k}} \tag{14}$$

$$\hat{\boldsymbol{\Sigma}}_{d,k} = \frac{\sum_{s=1}^{S_d} \hat{\theta}_{s,k} \{ \hat{\boldsymbol{\Sigma}}_s + (\hat{\boldsymbol{\mu}}_s - \hat{\boldsymbol{\mu}}_{d,k})(\hat{\boldsymbol{\mu}}_s - \hat{\boldsymbol{\mu}}_{d,k})^{\top} \}}{\sum_{s=1}^{S_d} \hat{\theta}_{s,k}} \tag{15}$$

これらトピック文の事後分布の平均から、各文書ごとにトピック文を生成する：$\hat{\boldsymbol{w}}_{d,k} \sim \mathrm{RNN}(\hat{\boldsymbol{\mu}}_{d,k})$。既存研究 [4] と同様に、潜在分布の平均を用いることで、当該トピックの要約文が得られることを期待する。

3.3 Gaussian word embedding との関連

親トピック文と子トピック文の事後分布間の KL ダイバージェンス（**KLD**）は、式 (12) に示した変分下限の $\boldsymbol{x}_s$ に関する項によって上から抑えられる（6.2 節にて証明）。

表 1 評価データセットにおける各モデルの ROUGE-F 値。

データセット	Yelp Dataset Challenge			Amazon Product Reviews		
評価指標	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L
Multi-Lead-1	27.42	3.74	14.34	30.32	**5.85**	15.96
LexRank [12]	26.53	3.30	14.54	31.42	5.17	16.67
Opinosis [13]	25.80	2.92	14.57	28.90	4.11	16.33
MeanSum [9]	28.66	3.73	15.77	30.16	4.51	17.76
Copycat [4]	28.95	4.80	**17.76**	**31.84**	5.79	**20.00**
提案法	**29.99**	**5.03**	17.39	31.31	5.64	18.03

表 2 各モデルにより生成された要約の人手評価スコア。

モデル	Fluency	Non-red.	Ref. Clarity	Coherence	Focus	Fluency	Non-red.	Ref. Clarity	Coherence	Focus
LexRank	-0.16	-0.15	-0.14	-0.06	-0.19	-0.14	-0.10	-0.16	-0.10	-0.10
MeanSum	-0.01	-0.06	-0.02	-0.01	-0.03	-0.05	-0.13	-0.04	-0.04	-0.06
Copycat	**0.09**	**0.16**	0.06	**0.04**	0.08	0.09	**0.12**	**0.13**	**0.11**	0.02
提案法	0.08	0.05	**0.09**	0.02	**0.14**	**0.10**	0.11	0.07	0.04	**0.14**

$$\begin{aligned}&\textstyle\sum_{s=1}^{S_d}\hat{\theta}_{s,k}\mathrm{D}_{\mathrm{KL}}\left[q(\boldsymbol{x}_s|\boldsymbol{w}_s)|p(\boldsymbol{x}_s|z_s=k)\right]\\ &\geq\textstyle\sum_{s=1}^{S_d}\hat{\theta}_{s,k}\mathrm{D}_{\mathrm{KL}}\left[q(\boldsymbol{x}_s|z_s=k)|p(\boldsymbol{x}_s|z_s=k)\right]\\ &=\textstyle\sum_{s=1}^{S_d}\hat{\theta}_{s,k}\mathrm{D}_{\mathrm{KL}}\left[q(\boldsymbol{x}_s|z_s=k)|q(\boldsymbol{x}_s|z_s=par(k))\right]\end{aligned}\quad(16)$$

一方、Gaussian word embedding では、単語のタクソノミー上において、親と子の潜在分布間の KLD を最小化することで、「動物」といった一般的な単語の分散は大きく、「犬」や「猫」などの具体的な単語の分散は小さくなることが報告されている [24]。これは「動物」は「犬」を指すこともあれば「猫」を指すこともあり、意味の分散が大きいと説明できる。本研究も同様に、変分下限の最大化により、親子の潜在分布間の KLD の上界が最小化され、親トピック文の分散は大きく、子トピック文の分散が小さくなる。例えば、“I love this restaurant.” など全体的な内容に関する文は、「料理」に関する内容を指すこともあれば、「サービス」に関する内容を指すこともあり、意味の分散が大きい。したがって、上位トピックほど潜在分布の分散が大きく、より一般的な文が生成されることが期待される。

3.4 トピック文の抽出

最後に、生成されたトピック文から要約に適切な文を抽出して要約を作成する。既存研究 [9] では、要約候補と入力文書間の ROUGE スコアは、参照要約との ROUGE スコアと高い相関があり、参照要約が存在しない状況下での要約候補評価に有用なことが知られている。本研究はこの知見を利用し、入力文書との ROUGE-1 (F1) を最大化するトピック文集合をビーム探索で選択することで、要約として相応しいトピックと詳細度合いを持つトピック文を参照要約なしに得る。その際に、既に要約に含まれている文との ROUGE-1 (precision) が高い文を追加しないことで、要約が冗長になることを防止する。

表 3 データセット中のレビュー文書/要約の件数。

データセット	Yelp	Amazon
学習	1,012,280/–	4,566,519/–
検証	800/100	672/84
評価	800/100	768/96

4. 実験

4.1 実験設定

本実験では、Yelp と Amazon のレビューを用いて複数文書要約生成の性能評価を行った。レビュー文書の件数を表 3 に示す。参照要約は Amazon Mechanical Turk にて、各製品について 8 件のレビューの要約を依頼し作成されたものである [4], [9]。ベースラインには、各文書の先頭の文を抽出する *Multi-Lead-1* に加え、教師なし抽出型手法である *LexRank* [12] と *Opinosis* [13]、及び教師なし生成型手法の *MeanSum* [9] と *Copycat* [4] を用いた。エンコーダ・デコーダには GRU-RNN を用い、単語の潜在表現・隠れ層・文の潜在表現の次元数はそれぞれ 200、400 及び 32 である。勾配降下法は Adam [16] を用い、学習率は 10^{-3} でバッチサイズは 4 である。木構造は 3 階層で固定され、それぞれの親トピックは 3 つの子トピックを持つ。

4.2 生成された要約の定量評価

表 1 に、評価データセットにおける各モデルの ROUGE スコアを示す。太字はモデル間で最大のスコアを示し、下線で示したスコアは approximate randomization test において、最大スコアとの差異が統計的に優位であるとは言えないことを示す（$p < 0.05$）。いずれのデータセットにおいても、ほぼ全ての評価指標において、提案法は MeanSum

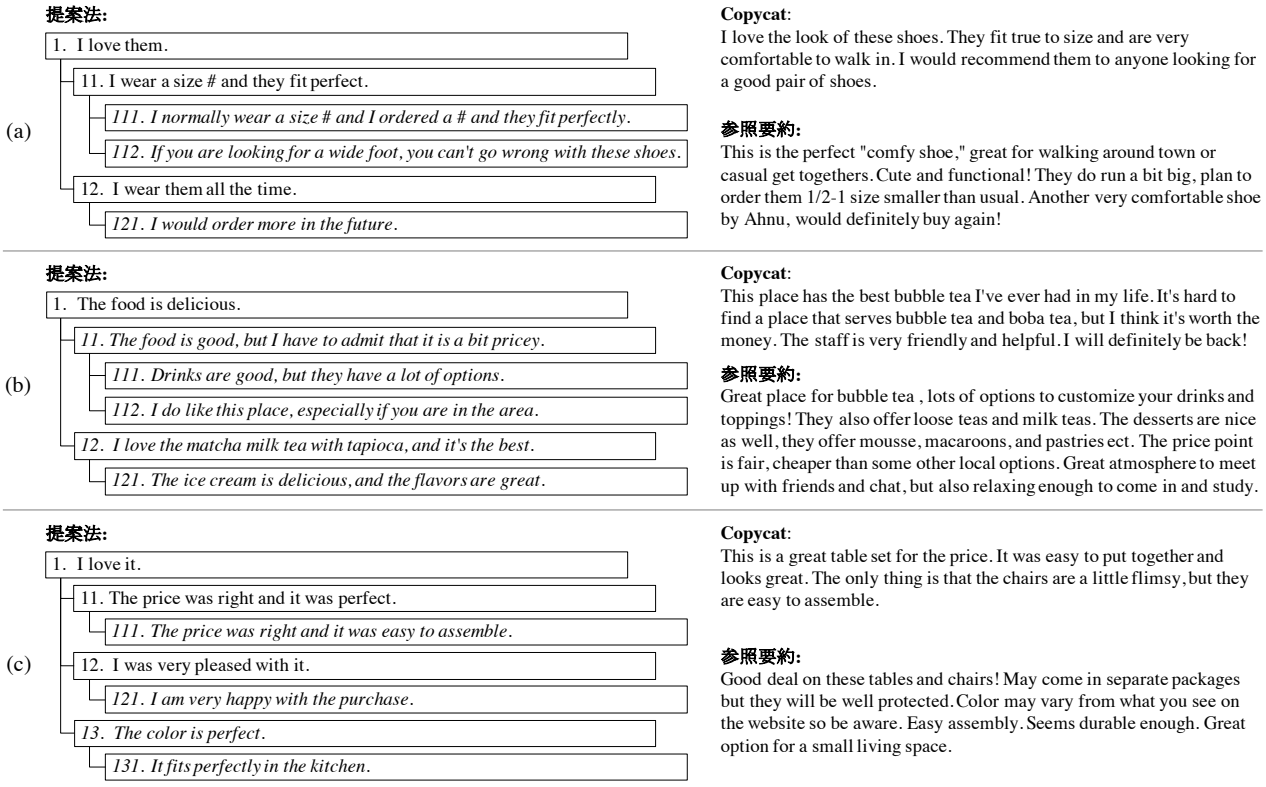


図 3 生成された要約と木構造の例。斜線は要約として選択されたトピック文を示す。

を上回り、最新の教師なし要約モデルである Copycat に対しては競合する性能が得られた。

次に、得られた要約を基準別に評価するために、Amazon Mechanical Turk 上にて人手評価を行った。既存研究 [1], [4] に基づき、各評価データセットから 50 件のレビューと要約のペアを無作為に選び、5 人の評価者に評価を依頼した。要約元の 8 つのレビューと、そのレビューに基づき各モデルにより生成された要約のペアを無作為な順序で提示し、要約の順位付けを依頼した。評価スコア算出では、最高と評価された回数から最低と評価された回数を差し引いた割合で各モデルの人手評価スコアを算出している。これは best-worst scaling [20] と呼ばれ、最高/最低と評価された回数のみを評価スコア算出に考慮することで、頑強性が高い評価が行えることが報告されている [17]。

評価基準は、既存研究 [9], [10] に基づき、以下の 5 つの基準を採用した。*Fluency*: 文法的に正しく、読みやすく、理解しやすい。*Non-redundancy (Non-red.)*: 不必要な反復的な単語やフレーズがない。*Referential-clarity (Ref.-clarity)*: 代名詞や名詞句が誰のことを指しているのか、何を指しているのかがわかる。*Coherence*: よく構成されていて整理されている。*Focus*: 各文について、要約中の他の文に関連する情報を含んでいる。

表 2 に、各モデルにより生成された要約の人手評価スコアを示す。−1 は全員が全てのペアについて最高と評価し、+1 は全員が全てのペアについて最低と評価した場合のスコアを示す。太字はモデル間で最大のスコアを示し、下線で示したスコアはテューキー・クレーマー検定において、最大スコアとの差異が統計的に優位とは言えないことを示す（$p < 0.05$）。両データセットを通じて、Copycat と提案法は他の 2 つの手法よりも高いスコアが得られた。提案法はいずれの基準においても Copycat とのスコアの差が統計的に有意とは言えず、Copycat と競合していることがわかる。評価基準を比較すると、提案法は特に Focus の観点で高い評価を得ており、トピックとその詳細度合いを明示的に考慮することで、特定のトピックに焦点を当てた要約を生成することができることが示された。

4.3 生成された要約の定性評価

本節では、生成された要約とその木構造の例を示しながら、提案法の長所と短所について議論する。

図 3 (a) に、靴に関する Amazon レビューから生成された要約を示した。提案法は、11 番目のトピック文とその子に記述されているように、「大きさ」や「履き心地」について述べており、12 番目のトピック文とその子は、レビュワーがその靴に満足していることについて言及している。Copycat もまた「大きさ」や「履き心地」について触れているもののその記述は簡潔であり、提案法は重要なトピックにより焦点を当てた要約を生成していることがわかる。

図 3 (b) はコーヒーショップの Yelp レビューから生成された要約を示している。提案法も CopyCat も “bubble tea

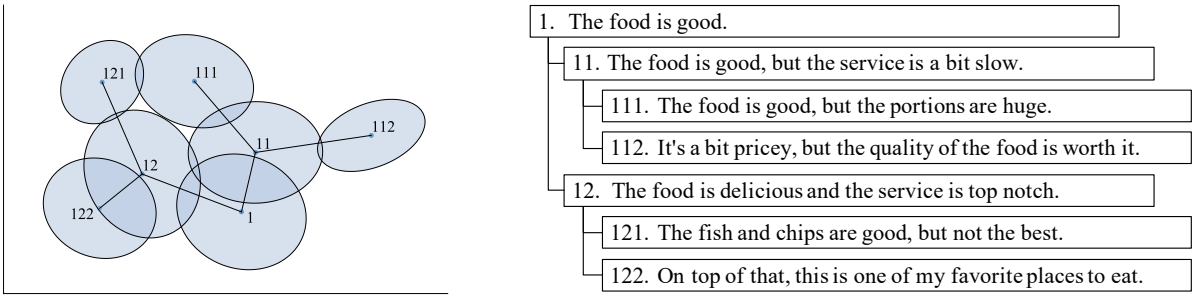


図 4 主成分分析による潜在空間から 2 次元空間への射影。各点はトピック文の潜在分布の平均を示しており、それぞれの円は平均からのマハラノビス距離が等しい座標群を示す。

表 4 トピック文の平均類似度（高いほど類似）。

親子間類似度	Yelp	Amazon
親子ペア	1.83	1.73
非親子ペア	1.25	1.24

表 5 トピック文の詳細度合い（高いほど詳細）。

階層毎の詳細度合い	Yelp	Amazon
第 1 階層	1.37	1.29
第 2 階層	1.67	1.45
第 3 階層	1.66	1.50

表 6 各階層における分散共分散行列の行列式の対数。

階層毎の $\log\lvert\hat{\mathbf{\Sigma}}_{d,k}\rvert$ の平均値	Yelp	Amazon
第 1 階層	21.5	19.0
第 2 階層	19.6	18.1
第 3 階層	15.6	13.3

(tea with tapioca)” の美味しさに言及しているが、提案法は参照要約と同様にデザートにも着目していることがわかる。このように生成された要約に問題はないものの、12 番目の文は 111 番目の文を詳述しており、葉に近くにつれトピック文がより具体的になるというモデル化の仮定に反していることがわかる。トピックとその木構造が固定されているために、木構造にこのような矛盾が生じることがあると考えられ、既存の木構造トピックモデル [15], [25] のように無限木構造を導入することで、より柔軟で一貫性のあるトピック木を生成できると期待される。

図 3 (c) に家具に関する Amazon レビューの要約を示す。提案法は価格、組み立ての簡単さ、見た目などの重要な要素を正確に捉えているものの、“table” や “chair” といった商品特有の情報を捉えることができていない。このような誤りは、商品の特性など商品特有の情報が多い Amazon レビューで特に散見された。既存研究 [4] では、注意機構やコピー機構 [22] により商品固有の情報を要約に取り込むことができることが報告されており、将来的にはこれら機構の導入が期待される。

4.4 生成されたトピック文とその木構造の定量分析

提案法で得られたトピック文と木構造について、親子間類似度と階層毎の詳細度合いの 2 つの観点から定量分析を行い、得られた木構造が仮説通りの特性を持つか検証した。

親子間類似度：親子関係にあるトピック文は類似するトピックについて述べていることが望ましい。この特性を確認するために、親子ペアと非親子ペアのトピック文の類似度を、semantic textual similarity タスク [8] で最高精度を達成している ALBERT [18] を用いて評価した。本実験では ALBERT-base を利用し、評価用ベンチマークでは、人手による類似度評価に対し 92.5%の相関係数が確認された。表 4 に示すように、いずれのデータセットにおいても、親子の関係性にあるトピック文は、親子でないトピック文のペアより類似しており、親子関係にあるトピック文は類似するトピックについて述べていることが確認された。

階層毎の詳細度合い：提案法では、階層が深くなるにつれ、トピック文の内容がより具体的なものになることを期待している。この特性を検証するために、文の詳細度合い推定タスク [19] にて ALBERT を学習し、各階層のトピック文の詳細度合いを計測した。評価用ベンチマークにおいて、人手による詳細度評価に対し 86.4%の相関係数が確認されており、これは当該タスクにおける最高精度である。表 5 に示すように、上位トピックからは一般的な、下位トピックからはより詳細な内容に関する文が生成されている。したがって、階層が深くなるにつれ、トピック文の内容がより具体的なものになることが確認された。

4.5 トピック文の潜在分布の分析

図 4 に、あるレストランレビューに関するトピック文の潜在分布について、その主成分ベクトル空間を可視化した。仮説通り子トピック文の潜在分布は親のトピック文の分布に近い位置にあり、親子関係にあるトピック文が類似したトピックを持つものと考えられる。また表 6 に示すように、各階層における分散共分散行列の行列式の対数は、階層が深くなるにつれ小さくなることが確認された。表 5 の結果も踏まえると、葉に近づくにつれ潜在分布の分散が小さくなり、具体的な文が生成されることがわかる。

Vol.2020-NL-246 No.31
Vol.2020-SLP-134 No.31
2020/12/3

5. おわりに

本研究では、文書から木構造上のトピックを推定し、トピック毎に要約文を生成することで、意見文書の要約が教師なしに得られることを示した。評価実験において、提案法は最新の教師なし生成型要約手法と競合する性能を有することが確認された。また、根の文の潜在分布は分散が大きく一般的な文が生成される一方、葉に近づくにつれ分散が小さくなり具体的な文が生成されるといった特性を確認した。これは単語の潜在表現にガウス分布を用いたGaussian word embeddingにて報告された特性と類似しており、要約のみならず、質問応答や対話生成などの文の詳細度合いを考慮する他タスクにも有用な知見である。

謝辞 本研究は、JST ACT-X JPMJAX1904、JST CREST JPMJCR1513及びJSPS特別研究員奨励費20J10726の支援を受けたものである。

6. 付録

6.1 木構造上のトピック分布の推定

本研究では、勾配降下法により学習可能な木構造ニューラルトピックモデル（TSNTM [15]）を用いることで、文の木構造上のトピック分布を推定する。

まず前置きとして、木構造トピックモデルの一つであるhierarchical Dirichlet process [21]は、各文に関する木構造上のパス分布 $\boldsymbol{\pi}_s$ と階層分布 $\boldsymbol{\phi}_s$ を下記式でモデル化する。

$$\nu_{s,k} \sim \mathrm{Beta}(1,\gamma),\ \pi_{s,k} = \pi_{s,par(k)}\nu_{s,k} \prod_{j\in \mathrm{Sib}(k)} (1-\nu_{s,j}) \quad (17)$$

$$\eta_{s,k} \sim \mathrm{Beta}(\alpha,\beta),\ \phi_{s,k} = \eta_{s,k} \prod_{j\in \mathrm{Anc}(k)} (1-\eta_{s,j}) \quad (18)$$

$$\theta_{s,k} = \pi_{s,k}\cdot\phi_{s,k} \quad (19)$$

ここで、$\mathrm{Sib}(k)$ 及び $\mathrm{Anc}(k)$ はそれぞれトピック k の前の兄弟のトピック集合及び祖先のトピック集合を指す。図5に示すように、$\pi_{s,k}$ は文 s が根トピックからトピック k を通るパスを選択する確率を示す。一方、$\phi_{s,k}$ は当該パス上で文 s がトピック k の祖先のトピック $j \in \mathrm{Anc}(k)$ を選択せず、トピック k を選択する確率を示す。文 s がトピック k を選択する確率 $\theta_{s,k}$ はこれら2つの確率の積で表される。

一方、TSNTMは、doubly-recurrent neural networks (DRNN) を用いて文の潜在表現 $\boldsymbol{y}_s = \mathrm{RNN}(\boldsymbol{w}_s)$ をパス分布 $\boldsymbol{\pi}_s$ 及び階層分布 $\boldsymbol{\phi}_s$ に変換する。DRNNは親子と兄弟間の2つのRNNデコーダで構成され、トピック k の中間層は式 (20) で表される。この中間層から得られた棒折比率 $\boldsymbol{\nu}_s$ で式を置き換え、パス分布を得る。

$$\boldsymbol{h}_k = \tanh(\boldsymbol{W}_p\boldsymbol{h}_{par(k)} + \boldsymbol{W}_s\boldsymbol{h}_{k-1}) \quad (20)$$

$$\nu_{s,k} = \mathrm{sigmoid}(\boldsymbol{h}_k^\top \boldsymbol{y}_s) \quad (21)$$

ただし $\boldsymbol{h}_{par(k)}$ 及び $\boldsymbol{h}_{k-1}$ はそれぞれトピック k の親とその前の兄弟である。同様に、階層分布もまたDRNNにより棒折比率 $\boldsymbol{\eta}_s$ を算出することで得られる。

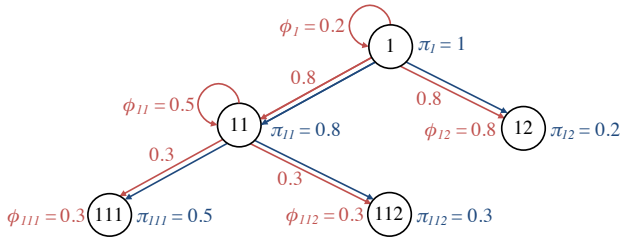


図 5 パス分布（青）と階層分布（赤）の例。パス分布の各階層における合計と、階層分布の各パスにおける合計は1になる。

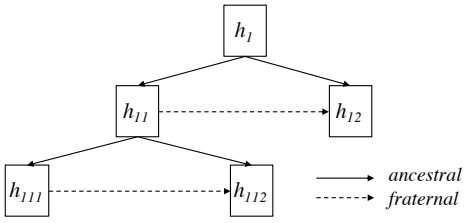


図 6 Doubly-recurrent neural networksの概要図。

6.2 式 (16) の導出

命題： $q(\boldsymbol{x}_s|z_s)$ が式 (13)、(14)、(15) により与えられる時、式 (22) が成り立つ:

$$\begin{aligned}&\textstyle\sum_s \hat{\theta}_{s,k}\mathrm{D}_{\mathrm{KL}}\left[q(\boldsymbol{x}_s|\boldsymbol{w}_s)|p(\boldsymbol{x}_s|z_s=k)\right]-\\ &\textstyle\sum_s \hat{\theta}_{s,k}\mathrm{D}_{\mathrm{KL}}\left[q(\boldsymbol{x}_s|z_s=k)|p(\boldsymbol{x}_s|z_s=k)\right] \geq 0\end{aligned} \quad (22)$$

証明： 式 (22) の第1項は以下のように展開できる。

$$\begin{aligned}&\textstyle\sum_s \hat{\theta}_{s,k}\mathrm{D}_{\mathrm{KL}}\left[q(\boldsymbol{x}_s|\boldsymbol{w}_s)|p(\boldsymbol{x}_s|z_s=k)\right]\\ =&\textstyle\sum_s \hat{\theta}_{s,k}\mathrm{D}_{\mathrm{KL}}\left[\mathcal{N}(\hat{\boldsymbol{\mu}}_s,\hat{\boldsymbol{\Sigma}}_s)|\mathcal{N}(\hat{\boldsymbol{\mu}}_{d,par(k)},\hat{\boldsymbol{\Sigma}}_{d,par(k)})\right]\\ =&\frac{1}{2}\textstyle\sum_s \hat{\theta}_{s,k}\Big\{\log|\hat{\boldsymbol{\Sigma}}_{d,par(k)}|-\log|\hat{\boldsymbol{\Sigma}}_s|+\mathrm{Tr}[\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}\hat{\boldsymbol{\Sigma}}_s]\\ &+(\hat{\boldsymbol{\mu}}_s-\hat{\boldsymbol{\mu}}_{d,par(k)})^\top\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}(\hat{\boldsymbol{\mu}}_s-\hat{\boldsymbol{\mu}}_{d,par(k)})-d\Big\}\\ =&\frac{1}{2}\textstyle\sum_s \hat{\theta}_{s,k}\Big\{C_{d,par(k)}-\log|\hat{\boldsymbol{\Sigma}}_s|+\mathrm{Tr}[\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}\hat{\boldsymbol{\Sigma}}_s]\\ &+\hat{\boldsymbol{\mu}}_s^\top\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}\hat{\boldsymbol{\mu}}_s-2\hat{\boldsymbol{\mu}}_{d,par(k)}^\top\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}\hat{\boldsymbol{\mu}}_s\Big\}\\ =&\frac{1}{2}\textstyle\sum_s \hat{\theta}_{s,k}\Big\{C_{d,par(k)}-\log|\hat{\boldsymbol{\Sigma}}_s|+\mathrm{Tr}[\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}\hat{\boldsymbol{\Sigma}}_s]\\ &+\mathrm{Tr}[\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}\hat{\boldsymbol{\mu}}_s\hat{\boldsymbol{\mu}}_s^\top]-2\hat{\boldsymbol{\mu}}_{d,par(k)}^\top\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}\hat{\boldsymbol{\mu}}_{d,k}\Big\}\end{aligned} \quad (23)$$

ただし、式 (14) から $\sum_s \hat{\theta}_{s,k}\hat{\boldsymbol{\mu}}_{d,k} = \sum_s \hat{\theta}_{s,k}\hat{\boldsymbol{\mu}}_s$ を用いた。

第2項についても同様に、以下のように展開できる。

$$\begin{aligned}&\textstyle\sum_s \hat{\theta}_{s,k}\mathrm{D}_{\mathrm{KL}}\left[q(\boldsymbol{x}_s|z_s=k)|p(\boldsymbol{x}_s|z_s=k)\right]\\ =&\frac{1}{2}\textstyle\sum_s \hat{\theta}_{s,k}\Big\{C_{d,par(k)}-\log|\hat{\boldsymbol{\Sigma}}_{d,k}|+\mathrm{Tr}[\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}\hat{\boldsymbol{\Sigma}}_{d,k}]\\ &+\mathrm{Tr}[\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}\hat{\boldsymbol{\mu}}_{d,k}\hat{\boldsymbol{\mu}}_{d,k}^\top]-2\hat{\boldsymbol{\mu}}_{d,par(k)}^\top\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}\hat{\boldsymbol{\mu}}_{d,k}\Big\}\end{aligned} \quad (24)$$

したがって、式 (22) は以下のように整理される。

$$\begin{aligned}&\textstyle\sum_s \hat{\theta}_{s,k}\mathrm{D}_{\mathrm{KL}}\left[q(\boldsymbol{x}_s|\boldsymbol{w}_s)|p(\boldsymbol{x}_s|z_s=k)\right]\\ &-\textstyle\sum_s \hat{\theta}_{s,k}\mathrm{D}_{\mathrm{KL}}\left[q(\boldsymbol{x}_s|z_s=k)|p(\boldsymbol{x}_s|z_s=k)\right]\\ =&\frac{1}{2}\textstyle\sum_s \hat{\theta}_{s,k}\Big\{-\log|\hat{\boldsymbol{\Sigma}}_s|+\log|\hat{\boldsymbol{\Sigma}}_{d,k}|\\ &+\mathrm{Tr}\left[\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}(\hat{\boldsymbol{\Sigma}}_s+\hat{\boldsymbol{\mu}}_s\hat{\boldsymbol{\mu}}_s^\top-\hat{\boldsymbol{\Sigma}}_{d,k}-\hat{\boldsymbol{\mu}}_{d,k}\hat{\boldsymbol{\mu}}_{d,k}^\top)\right]\Big\}\\ =&\frac{1}{2}\textstyle\sum_s \hat{\theta}_{s,k}\Big\{-\log|\hat{\boldsymbol{\Sigma}}_s|+\log|\hat{\boldsymbol{\Sigma}}_{d,k}|\Big\}+\frac{1}{2}\Big\{\mathrm{Tr}[\hat{\boldsymbol{\Sigma}}_{d,par(k)}^{-1}\\ &\textstyle\sum_s \hat{\theta}_{s,k}(\hat{\boldsymbol{\Sigma}}_s+\hat{\boldsymbol{\mu}}_s\hat{\boldsymbol{\mu}}_s^\top-\hat{\boldsymbol{\Sigma}}_{d,k}-\hat{\boldsymbol{\mu}}_{d,k}\hat{\boldsymbol{\mu}}_{d,k}^\top)]\Big\}\\ =&\frac{1}{2}\textstyle\sum_s \hat{\theta}_{s,k}\Big\{-\log|\hat{\boldsymbol{\Sigma}}_s|+\log|\hat{\boldsymbol{\Sigma}}_{d,k}|\Big\}\end{aligned} \quad (25)$$

ただし、式 (15) より $\sum_s \hat{\theta}_{s,k}\{\hat{\boldsymbol{\Sigma}}_s+\hat{\boldsymbol{\mu}}_s\hat{\boldsymbol{\mu}}_s^\top\} = \sum_s \hat{\theta}_{s,k}\{\hat{\boldsymbol{\Sigma}}_{d,k}+\hat{\boldsymbol{\mu}}_{d,k}\hat{\boldsymbol{\mu}}_{d,k}^\top\}$ を用いた。すなわち、与式はエントロピーの重み付け和の大小比較に帰着される。

一般に、$-\int q_1(\boldsymbol{x})\log q_2(\boldsymbol{x})\ d\boldsymbol{x} \geq -\int q_1(\boldsymbol{x})\log q_1(\boldsymbol{x})\ d\boldsymbol{x}$ が成立することから、式 (26) が成り立つ。

$$\begin{aligned}&\textstyle\sum_s \hat{\theta}_{s,k}\Big\{-\int q(\boldsymbol{x}_s|\boldsymbol{w}_s)\log q(\boldsymbol{x}_s|z_s=k)\ d\boldsymbol{x_s}\Big\}\\ &\geq\textstyle\sum_s \hat{\theta}_{s,k}\Big\{-\int q(\boldsymbol{x}_s|\boldsymbol{w}_s)\log q(\boldsymbol{x}_s|\boldsymbol{w}_s)\ d\boldsymbol{x_s}\Big\}\end{aligned} \tag{26}$$

右辺はガウス分布のエントロピーの重み付け和であることから、以下のように書ける。

$$\begin{aligned}&\textstyle\sum_s \hat{\theta}_{s,k}\Big\{-\int q(\boldsymbol{x}_s|\boldsymbol{w}_s)\log q(\boldsymbol{x}_s|\boldsymbol{w}_s)\ d\boldsymbol{x_s}\Big\}\\ =&\frac{1}{2}\textstyle\sum_s \hat{\theta}_{s,k}\Big\{\log|\hat{\boldsymbol{\Sigma}}_s|+d\log 2\pi+d\Big\}\end{aligned} \tag{27}$$

一方、左辺は以下のように展開できる。

$$\begin{aligned}&\textstyle\sum_s \hat{\theta}_{s,k}\Big\{-\int q(\boldsymbol{x}_s|z_s=k)\log q(\boldsymbol{x}_s|\boldsymbol{w}_s)\ d\boldsymbol{x_s}\Big\}\\ =&\frac{1}{2}\textstyle\sum_s \hat{\theta}_{s,k}\Big\{\log|\hat{\boldsymbol{\Sigma}}_{d,k}|+d\log 2\pi\\ &+\mathrm{E}_{q(\boldsymbol{x}_s|\boldsymbol{w}_s)}[(\boldsymbol{x}_s-\hat{\boldsymbol{\mu}}_{d,k})^\top\hat{\boldsymbol{\Sigma}}_{d,k}^{-1}(\boldsymbol{x}_s-\hat{\boldsymbol{\mu}}_{d,k})]\Big\}\end{aligned} \tag{28}$$

式 (28) の最後の項は以下のように整理できる。

$$\begin{aligned}&\textstyle\sum_s \hat{\theta}_{s,k}\Big\{\mathrm{E}_{q(\boldsymbol{x}_s|\boldsymbol{w}_s)}\big[(\boldsymbol{x}_s-\hat{\boldsymbol{\mu}}_{d,k})^\top\hat{\boldsymbol{\Sigma}}_{d,k}^{-1}(\boldsymbol{x}_s-\hat{\boldsymbol{\mu}}_{d,k})\big]\Big\}\\ =&\textstyle\sum_s \hat{\theta}_{s,k}\Big\{\mathrm{E}_{q(\boldsymbol{x}_s|\boldsymbol{w}_s)}[\mathrm{Tr}(\hat{\boldsymbol{\Sigma}}_{d,k}^{-1}(\boldsymbol{x}_s-\hat{\boldsymbol{\mu}}_{d,k})(\boldsymbol{x}_s-\hat{\boldsymbol{\mu}}_{d,k})^\top)]\Big\}\\ =&\mathrm{Tr}\Big[\hat{\boldsymbol{\Sigma}}_{d,k}^{-1}\textstyle\sum_s \hat{\theta}_{s,k}\{\mathrm{E}_{q(\boldsymbol{x}_s|\boldsymbol{w}_s)}[(\boldsymbol{x}_s-\hat{\boldsymbol{\mu}}_{d,k})(\boldsymbol{x}_s-\hat{\boldsymbol{\mu}}_{d,k})^\top]\}\Big]\\ =&\mathrm{Tr}\Big[\hat{\boldsymbol{\Sigma}}_{d,k}^{-1}\textstyle\sum_s \hat{\theta}_{s,k}\hat{\boldsymbol{\Sigma}}_{d,k}\Big]\\ =&\textstyle\sum_s \hat{\theta}_{s,k}\ d\end{aligned} \tag{29}$$

したがって、式 (26)、(27)、(28)、 (29) を整理すると、$\sum_s \hat{\theta}_{s,k}\big\{\log|\hat{\boldsymbol{\Sigma}}_{d,k}|\big\} \geq \sum_s \hat{\theta}_{s,k}\big\{\log|\hat{\boldsymbol{\Sigma}}_s|\big\}$ が成り立ち、命題が成立することが確認できる。

参考文献

[1] Amplayo, R. K. and Lapata, M.: Unsupervised Opinion Summarization with Noising and Denoising, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 1934–1945 (2020).

[2] Angelidis, S. and Lapata, M.: Summarizing Opinions: Aspect Extraction Meets Sentiment Prediction and They Are Both Weakly Supervised, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3675–3686 (2018).

[3] Bowman, S., Vilnis, L., Vinyals, O., Dai, A., Jozefowicz, R. and Bengio, S.: Generating Sentences from a Continuous Space, *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning*, pp. 10–21 (2016).

[4] Bražinskas, A., Lapata, M. and Titov, I.: Unsupervised Opinion Summarization as Copycat-Review Generation, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5151–5169 (2020).

[5] Carenini, G., Cheung, J. C. K. and Pauls, A.: Multi-document summarization of evaluative text, *Computational Intelligence*, Vol. 29, No. 4, pp. 545–576 (2013).

[6] Celikyilmaz, A. and Hakkani-Tur, D.: A hybrid hierarchical model for multi-document summarization, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 815–824 (2010).

[7] Celikyilmaz, A. and Hakkani-Tur, D.: Discovery of topically coherent sentences for extractive summarization, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 491–499 (2011).

[8] Cer, D., Diab, M., Agirre, E., Lopez-Gazpio, I. and Specia, L.: SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation, *Proceedings of the 11th International Workshop on Semantic Evaluation*, pp. 1–14 (2017).

[9] Chu, E. and Liu, P.: MeanSum: A Neural Model for Unsupervised Multi-Document Abstractive Summarization, *Proceedings of the 36th International Conference on Machine Learning*, pp. 1223–1232 (2019).

[10] Dang, H. T.: Overview of DUC 2005, *Proceedings of the Document Understanding Conference*, Vol. 2005, pp. 1–12 (2005).

[11] Dilokthanakul, N., Mediano, P. A., Garnelo, M., Lee, M. C., Salimbeni, H., Arulkumaran, K. and Shanahan, M.: Deep unsupervised clustering with gaussian mixture variational autoencoders, *CoRR*, Vol. arXiv:1611.02648v2 (2016).

[12] Erkan, G. and Radev, D. R.: LexPageRank: Prestige in Multi-Document Text Summarization, *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 365–371 (2004).

[13] Ganesan, K., Zhai, C. and Han, J.: Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions, *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 340–348 (2010).

[14] Gerani, S., Mehdad, Y., Carenini, G., Ng, R. T. and Nejat, B.: Abstractive summarization of product reviews using discourse structure, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1602–1613 (2014).

[15] Isonuma, M., Mori, J., Bollegala, D. and Sakata, I.: Tree-Structured Neural Topic Model, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 800–806 (2020).

[16] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, *CoRR*, Vol. arXiv:1412.6980v9 (2014).

[17] Kiritchenko, S. and Mohammad, S. M.: Capturing Reliable Fine-Grained Sentiment Associations by Crowdsourcing and Best–Worst Scaling, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 811–817 (2016).

[18] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P. and Soricut, R.: ALBERT: A Lite BERT for Self-supervised Learning of Language Representations, *Proceedings of the 7th International Conference on Learning Representations* (2019).

[19] Louis, A. and Nenkova, A.: Automatic identification of general and specific sentences by leveraging discourse annotations, *Proceedings of 5th International Joint Conference on Natural Language Processing*, pp. 605–613 (2011).

[20] Louviere, J. J., Flynn, T. N. and Marley, A. A. J.: *Best-worst scaling: Theory, methods and applications*, Cambridge University Press (2015).

[21] Paisley, J., Wang, C., Blei, D. M. and Jordan, M. I.: Nested hierarchical Dirichlet processes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

Vol. 37, No. 2, pp. 256–270 (2014).
[22] See, A., Liu, P. J. and Manning, C. D.: Get To The Point: Summarization with Pointer-Generator Networks, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1073–1083 (2017).
[23] Titov, I. and McDonald, R.: A joint model of text and aspect ratings for sentiment summarization, *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 308–316 (2008).
[24] Vilnis, L. and McCallum, A.: Word representations via gaussian embedding, *Proceedings of the 3rd International Conference on Learning Representations* (2015).
[25] Wang, C. and Blei, D. M.: Variational inference for the nested Chinese restaurant process, *Advances in Neural Information Processing Systems*, pp. 1990–1998 (2009).
[26] Wang, W., Gan, Z., Xu, H., Zhang, R., Wang, G., Shen, D., Chen, C. and Carin, L.: Topic-Guided Variational Auto-Encoder for Text Generation, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 166–177 (2019).
[27] Yang, Z., Hu, Z., Salakhutdinov, R. and Berg-Kirkpatrick, T.: Improved variational autoencoders for text modeling using dilated convolutions, *Proceedings of the 34th International Conference on Machine Learning*, pp. 3881–3890 (2017).