

指定語句を確実に含む見出し生成

山田 康輔^{1,2} 人見 雄太² 田森 秀明² 岡崎 直観³ 乾 健太郎^{4,5}

¹ 名古屋大学 ² 株式会社朝日新聞社 ³ 東京工業大学 ⁴ 東北大学 ⁵ 理化学研究所
yamada.kosuke@c.mbox.nagoya-u.ac.jp, processor.su2300@gmail.com,
tamori-h@asahi.com, okazaki@c.titech.ac.jp, inui@ecei.tohoku.ac.jp

1 はじめに

Rush ら [1] の研究以降、生成型要約モデルによる新聞記事の見出し生成が広く行われている。近年では、指定語句の挿入など、生成される見出しの制御に関する研究も盛んである。たとえば、Fan ら [2] は Convolutional Seq2Seq [3] を利用し、エンコーダに入力する記事の直前に指定語句を示す 1 トークンを挿入することで、指定語句を含んだ見出しを生成するよう制御した。また、He ら [4] は BART [5] を利用し、Fan らの手法の拡張として記事の直前に指定語句と “!” を挿入することで、複数トークンの語句指定を可能にした。加えて、生成結果への指定語句の出現率を高めるために学習時に指定語句を一定の確率でドロップアウトさせている。

しかし、これらの手法は生成された見出しが指定語句を含む確率を高めることは可能だが、指定語句を含む見出しが必ず生成されるとは限らない。記事を広告として用いるネイティブ広告向けの見出し生成では、見出しに指定語句を必ず含めなければならないケースがあり、実用では指定語句が確実に含まれるような見出し制御が必要となる。

そこで本研究では、指定語句を確実に含む見出しを生成する手法として、Dual Outward Generation mechanism (DOG) を提案する。DOG は見出しを生成する際に、文頭から文末へ生成するのではなく、指定語句から文頭および文末に向かって双方向に見出しを生成することで、指定語句を必ず含む見出し生成を実現する。表 1 に Fan ら [2] と He ら [4] の従来モデル¹および DOG を用い、語句を指定したときの見出し生成例を示す。また、図 1 に記事中の単語 1,2,3-gram を指定したとき、指定された語句が生成結果に含まれた割合 (指定語句出現率) を示す。

実験には日本語の新聞記事-見出しペアのコーパ

表 1: 従来モデルと DOG で語句を指定した見出し生成例

• 新聞記事: 刈谷市の伊勢湾岸自動車道のパーキングエリア「刈谷ハイウェイオアシス」の観覧車に、シースルーゴンドラが登場した。天井を除くイスや壁、床面が透明で、足元に広がる景色を 360 度楽しめる。観覧車は高さ 60 メートルで、1 周約 12 分。
• 人手で作成された見出し: 透明ゴンドラ登場 刈谷ハイウェイオアシスの観覧車
• 指定語句が「360 度楽しめる」の生成結果: DOG: 360 度楽しめる観覧車 刈谷ハイウェイオアシス He: 風に誘われ、360 度楽しむ 刈谷ハイウェイオアシス Fan: 360 度の観覧車、シースルーゴンドラ登場 刈谷ハイウェイオアシス
• 指定語句が「透明ゴンドラ登場」の生成結果: DOG: 観覧車、透明ゴンドラ登場 刈谷ハイウェイオアシス He: 観覧車にシースルーゴンドラ 刈谷 Fan: 透明な景色、360 度 刈谷ハイウェイオアシス

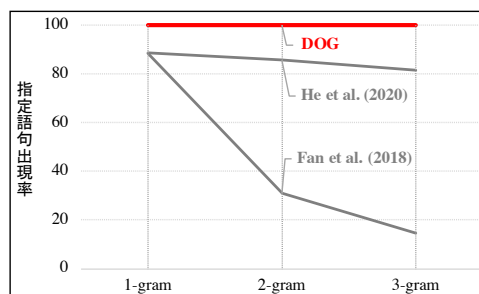


図 1: 記事中の単語の 1,2,3-gram を指定語句としたときの各モデルの生成結果の指定語句出現率

ス Japanese News Corpus (JNC) [7] の 2019 年版を利用し、指定語句出現率と自動要約の評価指標として代表的な ROUGE [8] にて提案手法を評価した。提案手法は指定語句出現率 100% を達成し、ROUGE の F 値も従来手法と同等なスコアを達成した。

2 提案手法

本研究では、指定語句を確実に含む見出し生成手法を提案する。具体的には、指定語句を $W (= y_{h:t})$ ²

² y_h と y_t はそれぞれ指定語句の先頭トークンと末尾トークンを示し、 $y_{h:t}$ はトークン列 $y_h \dots y_t$ を意味する。

¹ Seq2Seq モデルは Transformer [6] ベースに統一している。

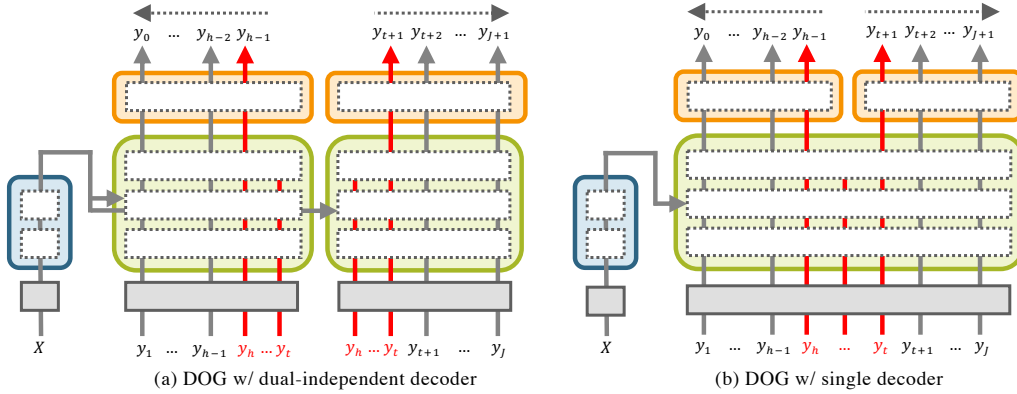


図 2: Dual Outward Generation mechanism (DOG) の概要図. 青枠が Transformer のエンコーダ, 緑枠がデコーダ, 橙枠が生成機構を示す. エンコーダからデコーダへの矢印は, エンコードされた隠れ状態をデコーダの注意機構で参照していることを指す. また, 点線の矢印は, 見出しの生成順序を示す.

として, 見出し $Y (= y_{0:h-1}, W, y_{t+1:j+1})$ を生成する場合, 指定語句の 1 つの前のトークン y_{h-1} から文頭トークン y_0 への生成と指定語句の 1 つ後のトークン y_{t+1} から文末トークン y_{j+1} への生成を同時に行うことで実現する. 提案モデルは Transformer [6] 型の Seq2Seq モデルを基本とする.

2.1 Transformer

Transformer [6] は注意機構をベースとしたエンコーダ-デコーダモデルである. エンコーダは自己注意機構と 2 層のフィードフォワードネットワーク (FFN) からなる Transformer Encoder 層を複数層積み上げることで構成され, デコーダは自己注意機構, 注意機構, 2 層の FFN からなる Transformer Decoder 層を複数層積み上げることで構成される. これらの入力は単語埋め込み層と位置エンコーディング層によって生成される記事と見出しの各単語の埋め込みを利用する. また, デコーダの出力から線形変換層と Softmax 層を通すことで見出しの各単語を生成する. 本研究では, これら 2 層を合わせて**生成機構**と定義する.

2.2 Dual Outward Generation mechanism

指定語句を確実に含む見出し生成する Dual Outward Generation mechanism (DOG) は指定語句から文頭へ向けた前向き生成と文末へ向けた後ろ向き生成を行うため, 生成機構を 2 つ持つ構造となる. この構造の場合, 前向き生成機構と後ろ向き生成機構で独立に生成するか, あるいは前向きと後ろ向き生成機構で生成過程を共有しつつ生成するか大きく 2 つのアイデアが考えられる.

2.2.1 各生成機構で独立に生成するモデル

前向きと後ろ向き生成機構で独立に生成するモデルは, 前向きと後ろ向きトークン列を個別に入力する必要があり, 各生成機構に対応する 2 つのデコーダを持つ. 見出し生成の各ステップでは, 前向きデコーダに指定語句と前向き生成トークン列, 後ろ向きデコーダに指定語句と後ろ向き生成トークン列を入力し, それぞれ独立に単語を生成する. 本モデルを DOG w/ dual-independent decoder と定義し, 図 2 (a) に概要を示す. また, 記事のトークン列 X , 見出しのトークン列 $Y (= y_{0:j+1})$, 指定語句のトークン列 $W (= y_{h:t})$, 指定語句から文頭までの見出しのトークン列 $\vec{Y} (= y_{0:h-1})$, 指定語句から文末までの見出しのトークン列 $\overleftarrow{Y} (= y_{t+1:j+1})$, \vec{Y} の長さ $L_{\vec{Y}} = h$, \overleftarrow{Y} の長さ $L_{\overleftarrow{Y}} = j - t + 1$ とし, 式 1 のように定式化される.

$$\begin{aligned}
 P(\vec{Y}|X, W) &= \prod_{j=1}^{L_{\vec{Y}}} P(y_{h-j}|y_{h-j+1:t}, X) \\
 P(\overleftarrow{Y}|X, W) &= \prod_{j=1}^{L_{\overleftarrow{Y}}} P(y_{t+j}|y_{h:t+j-1}, X)
 \end{aligned} \tag{1}$$

本モデルのデコーダの位置エンコーディング層に入力するトークン位置は, 前向きでは指定語句の末尾を 0 としたときの距離の値, 後ろ向きでは指定語句の先頭を 0 としたときの距離の値としている.

学習時には, 前向きと後ろ向き見出しを分けて学習させるため, 見出しを分割する必要がある.³そこで, 学習させる見出し中で無作為に 1 トークンを指定し, その位置を基準として見出しの分割を行う.

3 2.2.2 節のモデルも同様の処理が必要である.

2.2.2 生成機構間で生成過程を共有するモデル

2.2.1 節のモデルでは、前向きと後ろ向きでそれぞれ独立に生成しているため、見出し生成中の各ステップにおいて互いの生成機構が生成したトークンを考慮しておらず、前向きと後ろ向きで似たトークン列が生成される可能性がある。そこで、DOGの生成機構間で生成過程を共有するモデルを考える。具体的には、見出し生成中の各ステップにおいて、前向きの生成トークン列、指定語句のトークン列、後ろ向きの生成トークン列を連結して1つのトークン列にまとめて入力とする。ここではデコーダを2つ持つモデルと1つ持つモデルを検討する。デコーダが2つのモデルをDOG w/ dual decoder、デコーダが1つのモデルをDOG w/ single decoderと定義し、DOG w/ single decoderの概要を図2(b)に示す。また、式1と同様の記号を用いて、式2のように定式化される。

$$P(\vec{Y}, \vec{Y} | X, W) = \prod_{j=1}^{L_{max}} P(y_{h-j}, y_{t+j} | y_{h-j+1:t+j-1}, X) \quad (2)$$

$L_{max} = \max(L_{\vec{Y}}, L_{\vec{Y}})$ とし、 $y_{<0}$ と $y_{>J}$ はトークン列長を揃えるための特殊トークン“<pad>”とする。

本モデルのデコーダの位置エンコーディング層に入力するトークン位置は、中央から生成していくため、指定語句の中央を0とし、それより前向きは中央からの距離を負にした値、後ろ向きは中央からの距離の値としている。

また、本モデルの学習時には、デコーダの自己注意機構で使用するマスク方法に工夫を必要とする。通常のTransformerは文頭から文末に向けて見出しを生成するため、デコーダの自己注意機構にて入力時点以降の単語を参照しないために図3(a)に示すCausal Maskを活用している。しかし、DOGでは入力時点以降の単語を参照しないだけでなく、入力時点以前の単語も参照しないようにする必要がある。このため、図3(b)に示すDual Outward Causal Mask (DOC Mask) を利用している。また、本モデルは学習時の分割した見出しを前向きと後ろ向きで長さを揃える必要があることに注意する。

3 実験

3.1 実験設定

提案手法と従来手法を比較するために日本語の新聞記事から見出しを生成するためのコーパスで

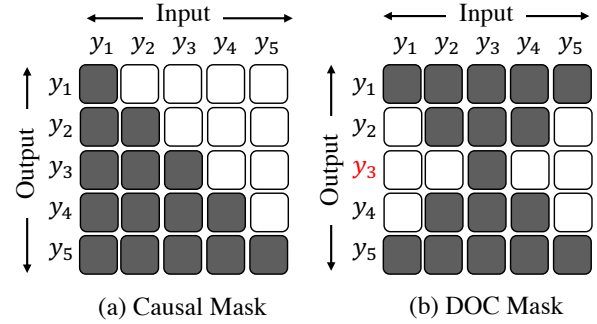


図3: Causal Mask (左)とDOC Mask (右)の概要図。白色のセルがマスクする箇所を示す。DOC Maskは学習時に y_3 を指定したときを示す。

あるJapanese News Corpus (JNC) [7]の2019年版を利用した。JNCには1,932,399件の記事-見出しペアが存在し、本研究ではこれらを学習セット、開発セット、テストセットとして98:1:1の比率で無作為に分割した。テキストの形態素解析にはipadic⁴を用いたMeCab [9]を使用し、Byte Pair Encoding (BPE)⁵ [10]によるサブワード化を行っている。

指定語句は人手で作成された見出し中の内容語とし、評価指標にはROUGE-1, ROUGE-2, ROUGE-Lの適合率、再現率、およびその調和平均であるF値を利用した。ROUGEによる評価する際には、人手で作成された見出しと生成された見出しから指定語句を除去している。これは従来手法では指定語句が必ず含まれた見出しになるとは限らず、本研究の提案手法では必ず指定語句が含まれた見出しとなっているため、指定語句を除去しなければ提案手法が有利な評価となってしまったためである。また、生成結果の長さを分析するために、生成された見出し数 n 、生成した見出しの長さを l_i 、人手で作成された見出しの長さを len としてAverage Length Error (ALE) ($= \frac{1}{n} \sum_{i=1}^n l_i - len$)を算出する。

比較する手法はTransformerをベースモデルとしたFanら[2]の手法とHeら[4]の手法とする。Fanらは記事の前に1トークンのみの指定語句を挿入、Heらは記事の前に複数トークンの指定語句と“!”を挿入し、学習時に指定語句を一定の確率でドロップアウトさせている。これらの手法の学習時の指定語句は人手で作成された見出しから抽出している。Heらの手法の学習時の指定語句は3-gramまでとする。

4 <https://taku910.github.io/mecab/>

5 <https://github.com/rsennrich/subword-nmt>

表 2: 従来手法および提案手法の実験結果. SR は指定語句出現率, P/R/F はそれぞれ適合率/再現率/F 値を示す.

	SR	ROUGE-1 P/R/F	ROUGE-2 P/R/F	ROUGE-L P/R/F	ALE	params ×10 ⁶
Transformer [6]	52.0	55.8/50.0/51.4	29.7/26.4/27.1	46.4/41.5/42.8	-3.3	72
Fan et al. (2018) [2] (Transformer Ver.)	94.0	57.7/ 51.2/52.7	31.1/ 27.4/28.2	48.0/ 42.4/43.9	-2.5	72
He et al. (2020) [4] (Transformer Ver.)	95.9	52.7/49.3/49.4	26.5/24.8/24.8	43.5/40.0/40.6	-1.2	72
DOG w/ dual-independent decoder	100.0	59.6/44.3/48.3	33.1/24.5/26.6	53.9/38.7/43.1	-5.8	106
DOG w/ dual decoder	100.0	60.3/47.5/51.5	32.9/25.8/27.9	50.9/40.1/43.6	-5.0	106
DOG w/ single decoder	100.0	59.3/47.9/51.4	31.9/25.6/27.5	50.0/40.2/43.3	-4.6	80

表 3: DOG w/ single decoder における生成例

新聞記事: ▽約 1 万匹のイワシが群れて泳ぐ様子を見られる京都水族館 (京都市)。この展示に合わせ、ちょっぴり変わった特別スイーツが 6 月末まで販売される。▽名前は「桜といわしのパフェ」で、桜でんぶと甘辛いイワシのショウガ煮の下には、抹茶とバニラのアイスや小豆……。			
人手で作成された見出し: 「目からウロコのおいしさ」 京都水族館にイワシパフェ			
指定語句	桜といわしのパフェ 6 月末まで 「目からウロコのおいしさ」 泳ぐ 展示	生成結果	「桜といわしのパフェ」 京都水族館で特別スイーツ イワシの特別スイーツ 京都水族館、6 月末まで販売 「目からウロコのおいしさ」 京都水族館で特別スイーツ イワシ泳ぐイワシ、特別スイーツに 京都水族館 展示

3.2 実験結果

表 2 に従来手法および提案手法の指定語句出現率 (SR), ROUGE-1, ROUGE-2, ROUGE-L の適合率, 再現率, F 値, ALE, パラメータ数 (params) を示す. 表 2 から従来手法が 95%前後の指定語句出現率なのに対し, 提案手法はモデルサイズを大きく変えずに 100%出現させることに成功している. また, 各 ROUGE の再現率は従来手法の方が高いが, 適合率は提案手法の方が高くなっており, F 値では同等のスコアとなっている. ここで, 提案手法で ROUGE の再現率が低いのは, ALE の値が小さいことから生成結果が全体的に短い傾向があるためであると考えられる. 以上の結果から, 指定語句を 100%含めた見出しを生成した上で, モデルサイズおよび ROUGE の F 値が同等である提案手法は, 実応用上においても有用であることが確認できる.

3.3 提案手法による見出し生成例

表 3 に DOG w/ single decoder による見出し生成例を示す. たとえば, 「桜といわしのパフェ」や「6 月末まで」のような商品名や日付情報を指定した結果, これらの語句に合わせて自然な見出しが生成できており, また, 記事中に存在しない「目からウロコのおいしさ」のようなフレーズを指定しても自然な生成ができていたことが確認できる. このような商品名や日付情報, フレーズの語句は忠実性が求められるため, 指定語句を確実に出現させる提案手法は実応用の観点で需要が高い. たとえば, プレスリ

リースを元にした記事や, ネイティブ広告の記事への見出し付与として活用が期待される.

しかし, DOG は両方向同時に生成するため, 「泳ぐ」を指定した生成例のように指定語句の前後どちらにも出現しうる単語を両方向共に生成させてしまったり, 「展示」を指定した生成例のように文頭と文末のどちらにも出現しやすいと判断された単語のときに, 両方向とも早い段階で生成を止める可能性があり, 課題が残る.

4 まとめと今後の展望

本研究では, 指定語句を確実に含んだ見出しを生成するために, 指定語句から文頭および文末に向かって見出しを生成する Dual Outward Generation mechanism (DOG) を提案した. JNC を用いて DOG を評価したところ, 従来手法では不可能だった指定語句出現率 100%の見出し生成を可能にし, その上で ROUGE の F 値にて同等のスコアを達成した.

実験では日本語の見出し生成を行ったが, DOG は多言語の見出し生成や機械翻訳などの他タスクにも応用可能だと考えている. DOG はモデルのデコーダ側の機構を変更したが, エンコーダ側で指定語句の位置情報などを取り込むことでスコアが改善されると考えている. また, ビームサーチなどの動的計画法による生成もスコアの改善が期待される.

謝辞

本研究は, 東京工業大学のスパコン TSUBAME3.0 を用いて行った.

参考文献

- [1] Alexander M. Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (ACL'15)*, pp. 379–389, 2015.
- [2] Angela Fan, David Grangier, and Michael Auli. Controllable abstractive summarization. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation (WNMT'18)*, pp. 45–54, 2018.
- [3] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. Convolutional sequence to sequence learning. In *Proceedings of the 34th International Conference on Machine Learning (ICML'17)*, pp. 1243–1252, 2017.
- [4] Junxian He, Wojciech Kryściński, Bryan McCann, Nazneen Rajani, and Caiming Xiong. CTRLsum: Towards generic controllable text summarization. *arXiv preprint arXiv:2012.04281*, 2020.
- [5] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL'20)*, pp. 7871–7880, 2020.
- [6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS'17)*, pp. 5998–6008, 2017.
- [7] Yuta Hitomi, Yuya Taguchi, Hideaki Tamori, Ko Kikuta, Jiro Nishitoba, Naoaki Okazaki, Kentaro Inui, and Manabu Okumura. A large-scale multi-length headline corpus for analyzing length-constrained headline generation model evaluation. In *Proceedings of the 12th International Conference on Natural Language Generation (INLG'19)*, pp. 333–343, 2019.
- [8] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out (WAS'04)*, pp. 74–81, 2004.
- [9] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to Japanese morphological analysis. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (ACL'04)*, pp. 230–237, 2004.
- [10] Philip Gage. A new algorithm for data compression. *C Users Journal*, Vol. 12, No. 2, pp. 23–38, 1994.