

テキスト生成におけるキーワード位置の制御

笹沢裕一 森下皓文 尾崎太亮 今一修 十河泰弘

日立製作所 研究開発グループ

{yuichi.sasazawa.bj, terufumi.morishita.wp, hiroaki.ozaki.yu,
osamu.imaichi.xc, yasuhiko.sogawa.tp}@hitachi.com

概要

テキスト生成における課題の1つは、ユーザの意図通りに生成を制御することである。既存研究ではキーワードを用いた制御が提案されているが、キーワードのみによる制御ではユーザの意図に沿ったテキストを生成するには不十分である。例えば、重要なキーワードをテキストの先頭近くに配置すれば読者の注意を引きやすくなるが、既存の手法ではそのような柔軟な制御はできない。本研究では各キーワードの位置も制御することで、ユーザの意図を効果的にテキストに反映できるようなテキスト生成に取り組む。タスクに依存しない汎用的な手法により、要約タスクと物語生成タスクにおいてキーワードとその位置を制御できることを示す。

1 導入

テキスト生成における課題の一つは、ユーザの意図に沿ったテキストを生成することである。多くの既存研究は、キーワードを利用してテキスト生成を制御する方法を提案している。テキスト要約では、要約テキストが含むべきキーワードをモデルに与えることで、原文書の特定の部分に焦点を当てたテキストを生成することができる [1, 2, 3]。物語生成では、キーワードと短いプロンプトを使用して、物語のストーリーを制御する [4, 5, 6]。その他の電子商取引生成、レビュー生成、質問生成などのタスクにおいても、キーワードを用いることでユーザにとって望ましいテキストを作成することができる [7, 8, 9, 10, 11]。

しかし、キーワードのみによる制御ではユーザの意図に沿ったテキストを生成するには不十分である。ユーザの意図をより反映させる方法としてキーワードの位置を制御する方法がある。例えば、重要なキーワードは読者の注意を引くためにテキストの先頭近くに配置し、一方で比較的重要でないがテキ

ストに入れる必要があるキーワードはテキストの後半に配置することができる。また、複数のキーワードの順番を指定できれば話の流れを制御することもできる。既存手法は単にキーワードを指定するのみであるためこのような柔軟な制御はできず、必ずしもユーザの意図するテキストが生成されるとは限らない。一部の既存研究はキーワードの順序の制御に取り組んでいるが [8, 12]、キーワード位置を直接制御する場合と比較して欠点が存在する。(a) キーワードの順序を制御するためには必ず複数のキーワードを指定する必要がある、単一のキーワードの位置を制御することができない。(b) 順序制御のみでは個々のキーワードの具体的な配置位置を必ずしも制御できない。例えば、指定した全てのキーワードがテキストの後半に配置されることがある。

本研究では、キーワードと各キーワードの位置を制御するテキスト生成に取り組む。既存研究ではテキスト生成タスクにおいてテキストスタイル、テキスト簡易化の種類、テキスト長などのテキスト属性を制御するために特殊トークンの使用が提案されている [13, 14, 15]。この手法をキーワード位置の制御に応用し、キーワードと各キーワードの相対位置を表す特殊トークンをモデルに与えることで、キーワード位置制御を行う。このシンプルな手法により、各キーワードを指定した位置に配置することが可能となる。また、位置制御の性能を向上させ、よりユーザの意図に沿ったテキストを生成するために、ターゲットテキストの長さを示す特殊トークンもモデルに与えて制御を行う。

要約タスクと物語生成タスクで評価を行い、このタスクに依存しない汎用的な手法によってキーワードとその位置を制御できることを示す (Section 3.2)。また、本手法が通常手法と比較してターゲット要約に近いテキスト、つまりユーザの意図に近いテキストを生成できることを示す (Section 3.3)。また、キーワードの位置を制御することで特定のキーワー

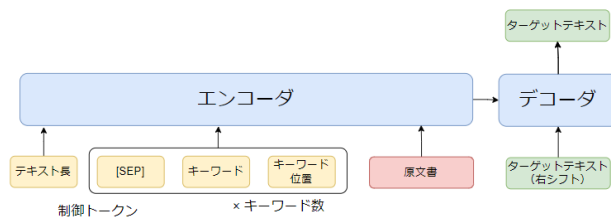


図1 提案手法の概要図。モデルにはテキスト生成を制御するための制御トークン：キーワード、各キーワードの位置、テキスト長が与えられる。

ドから多様なテキストを生成できることをケーススタディによって示す (Section 3.4).

2 手法

2.1 概要

要約タスクでは BART モデル [16] を、物語生成タスクでは GPT モデル [17] を使用する。テキスト生成を制御するために制御トークン：(1) ターゲットテキスト中のキーワード、(2) 各キーワードの位置、(3) ターゲットテキストの長さ、をモデルに与える。これらの制御トークンはターゲットテキストから獲得する。BART モデルを用いる場合は図 1 に示すように制御トークンを原文書と連結してエンコーダに与える。GPT モデルを用いる場合は制御トークンをデコーダに与える。BART モデルや GPT モデルを用いる通常のテキスト生成と同様に、モデルはクロスエントロピー損失を用いて条件付き確率 $p(y_i | y_{<i}, x)$ を最大化するように学習する。ここで y はターゲットテキスト、 x はモデルへの入力（制御トークンや要約タスクの原文書）である。

特殊トークンでテキストの属性を制御する既存の研究 [13, 14, 15] を応用し、各キーワードの位置とテキストの長さを特殊トークンとしてモデルに与える。例えば、キーワード “two dogs” がターゲットテキストの 20-30% の位置にあり、テキストの長さが 50-54 語の場合、モデルには “[LENGTH50][SEP]two dogs[POSITION20]” を制御トークンとして与える。

本研究では、学習時と推論時の両方においてターゲットテキストから獲得されたオラクル情報をモデルに与えることに留意されたい。本研究の目的はモデルに追加情報を与えることで意図したテキストを生成することであるため、この実験設定は適切なものである。キーワードとその位置をモデルが自動的に決定する（つまり、モデルにオラクル情報を与え

ない）設定も考えられるが、今後の課題としたい。

2.2 制御トークンの獲得

以下に記載する方法で、ターゲットテキストから制御トークンを獲得する。詳細は Appendix B に記載する。

キーワード ターゲットテキストからランダムにキーワードを抽出する。本研究でのキーワードは重要な単語に限定されず、ターゲットテキスト中の 1-3 個の連続した単語で構成されるフレーズを意味する。例えば、“Marcia was looking forward to try hang gliding.” というテキストから、“Marsha”, “was”, “looking forward”, “to trying”, “trying hang gliding” などのフレーズをキーワード候補として抽出する。ただし、“was” や “to trying” などの意味の薄い単語はユーザがキーワードとして与える可能性が低いと考えられるためキーワード候補から除外する。学習時には、キーワード候補の中からランダムな数のフレーズをキーワードとしてモデルに与える。

キーワード位置 各キーワードの位置を相対位置で表現する。具体的には、ターゲットテキストの先頭から数えたときの各キーワードの絶対位置をテキストの単語数で割り、10%単位で量子化した数値をモデルに与える。絶対位置ではなく相対位置を用いる理由は、ユーザがキーワードの具体的な絶対位置を指定したい場面は少なく、テキストの先頭付近、中央付近、末尾付近のような相対的な位置を指定する方が実用的だと考えられるためである。

テキストの長さ キーワードの相対位置の指定のみでは、モデルがキーワードの具体的な配置位置を決定することは難しいと考えられる。キーワードの相対位置を適切に制御するためにテキストの長さも同時に制御する。ユーザがキーワードの位置を制御する際にテキストの長さも同時に制御したいと考えるのは自然だと考えられる。我々はターゲットテキストの単語数を 5 語単位で量子化した数値をモデルに与える。

3 実験

3.1 実験設定

要約タスクと物語生成タスクにおいて評価を行う。この 2 つのタスクは異なる特徴を持つ。(1) 要約タスクでは、与えられた制御トークンに基づいて原文書から情報を抽出して短いテキストに圧縮す

表1 キーワードとその位置の制御の評価. 全てのキーワードを含むテキストが生成できた割合 (**包含**) と, 全てのキーワードをターゲットの位置に配置したテキストが生成できた割合 (**位置**) を評価する.

制御項目	CNN/DM		XSum		Stories	
	包含	位置	包含	位置	包含	位置
1つのキーワード						
キーワード無し	27.5	8.3	23.4	9.4	0.5	0.1
キーワード	71.3	18.7	86.4	28.7	53.0	14.3
+長さ	72.7	20.4	85.8	30.8	50.9	13.5
+位置	80.8	47.0	92.1	63.0	57.2	27.4
+位置+長さ	85.8	48.8	91.8	64.1	58.8	29.1
2つのキーワード						
キーワード	52.4	5.1	74.1	14.1	22.9	1.6
+位置+長さ	75.9	28.6	85.9	46.4	31.1	7.9
3つのキーワード						
キーワード	39.1	2.0	62.5	9.8	9.2	0.3
+位置+長さ	70.6	21.8	80.5	37.3	15.5	2.2

る。(2) 物語生成タスクでは, 与えられた制御トークンのみに基づいてテキストを生成する. 要約タスクでは CNN/DailyMail [18], XSum [19] データセットと BART_{LARGE} モデル (400M params) [16] を使用した. 物語生成タスクでは ROCStories [20] データセットと GPT2 モデル (120M params) [17] を使用した. 全ての実験において訓練と生成を3回行い, その平均スコアを報告する. より詳細については Appendix A を参照されたい.

3.2 キーワードの位置制御の評価

まず, キーワードがターゲットの位置に配置されているかを確認する. 具体的には, 全てのターゲットのキーワードを含むテキストが生成された割合と, 全てのキーワードが各ターゲット位置に配置したテキストが生成された割合を評価する.

表1は, 特殊トークンを使用する提案手法 (+位置, +位置+長さ) が, キーワードをターゲットの位置に含むテキストを生成できることを示している. また, キーワード位置と共にテキスト長を与える (+位置+長さ) ことで, 特にテキストが長いデータセット (CNN/DM, ROCStories) において位置制御の精度が向上している. この理由は, 相対位置と長さの情報を組み合わせることでモデルがキーワードを適切な位置に配置できるようになるためである. キーワードの位置が与えられると, キーワードの包含の精度も向上する. キーワードを配置すべき位置をあらかじめモデルに知らせることで, モデルが

キーワードをテキストに配置し忘れることを防いでいるためだと推測される. 物語生成タスクでは要約タスクに比べ制御精度が大幅に低下していることが分かる. 物語生成タスクでは原文書が与えられず制御トークンのみからテキストを生成するため, キーワードを含めるのに不適切な文脈を生成しやすいためと考えられる.

表2では, ターゲットのキーワード位置ごとにキーワードが正しい位置にあるテキストが生成されたかどうか, 生成されなかった場合はどの程度キーワードの位置がずれたか, またはキーワードが含まれなかったかを分析している. 全てのターゲット位置において, キーワードのみの制御を行った場合と比較してキーワード位置制御の精度が向上しており, 提案手法の有効性が示されている. また, テキストの先頭に近いほどキーワードの包含精度及び位置制御の精度が高いことが確認できる.

3.3 要約内容の制御の評価

テキストを制御することで, 要約タスクにおいてユーザが意図したテキストを生成することが容易になることを示す. 表4に示すように, ターゲットキーワードを与えることで ROUGE スコア [21] が通常の要約に比べて大きく向上していることがわかる. さらに, キーワードの位置とテキストの長さを制御することでスコアが向上していることがわかる. この結果により, キーワードの位置情報をモデルに与えることでユーザの意図する内容に近いテキストを生成しやすくなることが分かる.

3.4 ケーススタディ

提案手法によって生成された要約例を表3に示す. キーワードの位置が指定した位置からわずかにずれる場合も存在するが, 提案手法によってキーワードとその位置を制御できることが確認できる. また, キーワードの位置を変えることで, 複数の異なるテキストを生成されており, 特に各テキストのスタイルが異なることが分かる. 例えば, キーワード “true miracle dog” を0%の位置に配置すると, 冒頭の特徴的なキーワードで読者の注意を引くテキストが生成される. 一方, 90%の位置にキーワードを配置すると時系列で出来事を説明する物語調のテキストが生成される. 複数のキーワードを与えた場合においても, 各キーワードの位置を制御することでキーワードの順序を制御できることがわかる.

表 2 CNN/DM データセットにおけるキーワードとその位置の制御の詳細な評価。ターゲットの相対位置ごとに以下の 4 種類に分類する：キーワードがターゲットの位置に正しく配置されたか、ターゲット位置と実際の位置のずれが 10%以内か、位置のずれが 10%より大きいのか、もしくはキーワードが含まれていないテキストが生成されたか。

生成要約文での キーワード位置	ターゲットのキーワードの位置 (相対位置)									
	0-10%	10-20%	20-30%	30-40%	40-50%	50-60%	60-70%	70-80%	80-90%	90-100%
キーワードのみの制御										
正しい位置	52.6	23.8	14.5	9.5	9.5	9.1	8.7	11.8	12.7	15.6
キーワード+位置+文長の制御										
正しい位置	84.0	57.9	49.1	41.4	36.0	36.2	33.7	36.0	46.2	47.9
10%以下のずれ	8.1	27.5	31.9	34.4	36.1	34.1	35.5	34.3	23.3	8.9
10%より大きいずれ	3.2	5.3	8.3	12.8	15.1	15.5	15.1	11.4	6.7	10.9
キーワードを含まない	4.7	9.4	10.7	11.4	12.8	14.1	15.7	18.4	23.7	32.4

表 3 CNN/DM データセットにおける生成した要約例

キーワードとその位置	生成要約
true miracle dog (0%)	"She's a true miracle dog and she deserves a good life," foster mother says. Theia was apparently hit by a car and buried in a field. Four days later, she was found emaciated and dirt-covered by a farm worker. A fundraising page has raised more than \$10,000.
true miracle dog (40%)	A stray dog apparently hit by a car and buried in a field survives. The dog, named Theia, is "a true miracle dog ," her foster owner says. She suffered a dislocated jaw, leg injuries and a caved-in sinus cavity. A fundraising page has raised more than \$10,000.
true miracle dog (90%)	Stray dog apparently hit by a car and buried in a field survives. Four days later, she is found emaciated and dirt-covered on a farm. She has a dislocated jaw, leg injuries and a caved-in sinus cavity. A good Samaritan has set up a fundraising page for "the true miracle dog "
hit by (40%) bully breed mix (60%)	A stray dog in Washington State apparently died after being hit by a car. Four days later, she was found emaciated and dirt-covered in a field. Theia, a bully breed mix , has a dislocated jaw, leg injuries and a caved-in sinus cavity.
hit by (10%) bully breed mix (0%)	Theia, a bully breed mix , was apparently hit by a car and buried in a field. Four days later, she was found emaciated and dirt-covered by a farm worker. She suffered a dislocated jaw, leg injuries and a caved-in sinus cavity. A fundraising page has raised more than \$10,000.
キーワードとその位置	生成要約 (エラーケース)
buried in (80%)	Stray dog apparently hit by a car, apparently killed and buried in a field. Four days later, the dog manages to stagger to a nearby farm. She was found by a farm worker, who took her to a vet for help. Theia is now receiving treatment.
whacked (70%)	A stray dog in Washington State apparently died after being hit by a car. Four days later, she was found emaciated and dirt-covered by a farm worker. She suffered a dislocated jaw, leg injuries and a caved-in sinus cavity. A fundraising page has raised more than \$10,000.

表 4 ROUGE スコアを用いたテキスト要約の評価。本研究ではターゲットテキストから獲得したオラクル情報をモデルに与えているため、通常の要約と比較してスコアが大きく向上していることに留意されたい。

制御項目	CNN/DM			XSum		
	R1	R2	RL	R1	R2	RL
キーワード無し	44.3	21.3	41.2	45.8	22.6	37.6
キーワード	46.0	23.2	42.9	50.4	28.0	42.2
+長さ	47.4	24.1	44.3	51.4	28.8	43.2
+位置	46.6	23.9	43.6	51.3	29.1	44.7
+位置+長さ	48.3	25.1	45.2	52.4	30.0	45.7

さらにエラーケースを示す。テキストの末尾付近の位置を指定した場合、その指示を無視して全く別の位置にキーワードが配置されることや、テキストにキーワードが含まれないことが多い。表 2 の結果から確認できるように、提案手法はキーワードをテキストの後半に配置するのが苦手な傾向がある。理由の一つとして、テキストの末尾に近いほど指定したキーワードを含めるのに不適切な文脈になる可能

性が高まるためだと考えられる。

4 結論

本論文ではキーワードと各キーワードの位置を制御したテキスト生成に取り組んだ。要約タスクと物語生成タスクで実験を行い、タスクに依存しない手法によってキーワードの位置を制御できることを示した。また、本手法によってターゲットの要約に近いテキストが生成することが可能となり、ユーザの意図に近いテキストが生成できることを示した。今後の課題は、制御精度の向上及び、オラクル情報を使用せずに適切なキーワードを適切な位置に自動で配置することである。

謝辞

ご助言を頂いた、日立製作所研究開発グループの山口篤季氏、森尾学氏、是枝祐太氏に感謝を申し上げます。

参考文献

- [1] Angela Fan, David Grangier, and Michael Auli. Controllable abstractive summarization. In **Proceedings of the 2nd Workshop on Neural Machine Translation and Generation**, pp. 45–54, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [2] Junxian He, Wojciech Kryscinski, Bryan McCann, Nazneen Fatema Rajani, and Caiming Xiong. Ctrlsum: Towards generic controllable text summarization. **CoRR**, Vol. abs/2012.04281, , 2020.
- [3] Zi-Yi Dou, Pengfei Liu, Hiroaki Hayashi, Zhengbao Jiang, and Graham Neubig. GSum: A general framework for guided neural abstractive summarization. In **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 4830–4842, Online, June 2021. Association for Computational Linguistics.
- [4] Parag Jain, Priyanka Agrawal, Abhijit Mishra, Mohak Sukhwani, Anirban Laha, and Karthik Sankaranarayanan. Story generation from sequence of independent short descriptions. **CoRR**, Vol. abs/1707.05501, , 2017.
- [5] Angela Fan, Mike Lewis, and Yann Dauphin. Strategies for structuring story generation. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**, pp. 2650–2660, Florence, Italy, July 2019. Association for Computational Linguistics.
- [6] Lili Yao, Nanyun Peng, Ralph Weischedel, Kevin Knight, Dongyan Zhao, and Rui Yan. Plan-and-write: Towards better automatic storytelling. In **Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence**, AAAI’19/IAAI’19/EAAI’19. AAAI Press, 2019.
- [7] Zhangming Chan, Xiuying Chen, Yongliang Wang, Juntao Li, Zhiqiang Zhang, Kun Gai, Dongyan Zhao, and Rui Yan. Stick to the facts: Learning towards a fidelity-oriented E-commerce product description generation. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 4959–4968, Hong Kong, China, November 2019. Association for Computational Linguistics.
- [8] Huajie Shao, Jun Wang, Haohong Lin, Xuezhou Zhang, Aston Zhang, Heng Ji, and Tarek Abdelzaher. Controllable and diverse text generation in e-commerce. In **Proceedings of the Web Conference 2021**, WWW ’21, p. 2392–2401, New York, NY, USA, 2021. Association for Computing Machinery.
- [9] Jianmo Ni and Julian McAuley. Personalized review generation by expanding phrases and attending on aspect-aware representations. In **Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)**, pp. 706–711, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [10] Ying-Hong Chan, Ho-Lam Chung, and Yao-Chung Fan. Improving controllability of educational question generation by keyword provision. **CoRR**, Vol. abs/2112.01012, , 2021.
- [11] Zhiling Zhang and Kenny Zhu. Diverse and specific clarification question generation with keywords. In **Proceedings of the Web Conference 2021**, WWW ’21, p. 3501–3511, New York, NY, USA, 2021. Association for Computing Machinery.
- [12] Yixuan Su, David Vandyke, Sihui Wang, Yimai Fang, and Nigel Collier. Plan-then-generate: Controlled data-to-text generation via planning. In **Findings of the Association for Computational Linguistics: EMNLP 2021**, pp. 895–909, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [13] Kango Iwama and Yoshinobu Kano. Multiple news headlines generation using page metadata. In **Proceedings of the 12th International Conference on Natural Language Generation**, pp. 101–105, Tokyo, Japan, October–November 2019. Association for Computational Linguistics.
- [14] Surafel Melaku Lakew, Mattia Di Gangi, and Marcello Federico. Controlling the output length of neural machine translation. In **Proceedings of the 16th International Conference on Spoken Language Translation**, Hong Kong, November 2–3 2019. Association for Computational Linguistics.
- [15] Louis Martin, Éric de la Clergerie, Benoît Sagot, and Antoine Bordes. Controllable sentence simplification. In **Proceedings of the 12th Language Resources and Evaluation Conference**, pp. 4689–4698, Marseille, France, May 2020. European Language Resources Association.
- [16] 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**, pp. 7871–7880, Online, July 2020. Association for Computational Linguistics.
- [17] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2018.
- [18] Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, **Advances in Neural Information Processing Systems**, Vol. 28. Curran Associates, Inc., 2015.
- [19] Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Don’t give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, pp. 1797–1807, Brussels, Belgium, October–November 2018. Association for Computational Linguistics.
- [20] Nasrin Mostafazadeh, Nathanael Chambers, Xiaodong He, Devi Parikh, Dhruv Batra, Lucy Vanderwende, Pushmeet Kohli, and James Allen. A corpus and cloze evaluation for deeper understanding of commonsense stories. In **Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 839–849, San Diego, California, June 2016. Association for Computational Linguistics.
- [21] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In **Text Summarization Branches Out**, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [22] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, **3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings**, 2015.
- [23] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, pp. 2818–2826, 2016.
- [24] Yaoming Zhu, Sidi Lu, Lei Zheng, Jiaxian Guo, Weinan Zhang, Jun Wang, and Yong Yu. Tegygen: A benchmarking platform for text generation models. In **The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR ’18**, p. 1097–1100, New York, NY, USA, 2018. Association for Computing Machinery.

表5 データセットの統計

データ	#訓練	#開発	#評価	#原文書長	#テキスト長
CNN/DM	287,227	13,368	11,490	777.6	57.9
XSum	204,045	11,332	11,334	433.1	23.2
Stories	78,528	9,816	9,817	-	49.8

A 詳細な実験設定

表5に各データセットの訓練・開発・検証データのデータ数と原文書の単語数の平均値、ターゲットテキストの単語数の平均値を示す。オプティマイザはAdam [22]を使用し、 $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-6}$, L2正規化係数を0.01とした。学習率は訓練ステップの最初の6%でウォームアップさせ、その後線形に減衰させた。ドロップアウト率は0.1、バッチサイズは32、label smoothing [23]は0.1である。

要約タスク 学習率は 2×10^{-5} とした。ただし、新たに追加した特殊トークンの重みは初期化状態から学習させるため、単語埋め込みの重みに対する学習率は 1×10^{-3} と通常より高い値を使用した。エポック数は10回である。生成にはビームサーチを用いた。CNN/DMではビーム数は4、length penaltyは2.0、XSumではビーム数は6、length penaltyは1.0に設定した。原文書の最大トークン数を1024、要約文の最大トークン数を128とし、それ以上のトークン数のテキストは末尾を切り捨てた。

物語生成タスク 学習率は 2×10^{-5} 、単語埋め込みの重みの学習率は 1×10^{-3} 、エポック数は30回とした。生成には $p = 0.95$ のtop-pサンプリングを使用し、temperatureは0.1とした。ターゲットテキストの最大トークン数は128とし、それ以上のトークン数のテキストは末尾を切り捨てた。

B 制御トークン

テキスト長、キーワード、キーワード位置を獲得するための単語分割はNLTKライブラリを用いる。モデルの訓練時には、全ての制御トークン（キーワード、キーワード位置、テキスト長）をそれぞれ一定確率でモデルに与えて訓練し、この学習済みモデルを用いて4種類の設定（表1における「キーワード」、「+長さ」、「+位置」、「+位置+長さ」）でテキスト生成を行う。これにより、1種類のモデルで複数の設定での推論を行うことができ、設定ごとにモデルを訓練する場合と比較してコストを抑えて実験を行うことができる。予備実験の結果、各推論設

表6 Self-BLEUを用いた生成テキストの多様性の評価

制御項目	CNN/DM	XSum	Stories
	S-BLEU ↓	S-BLEU ↓	S-BLEU ↓
キーワード無し	96.2	92.2	-
キーワード	96.6	92.8	71.2
+位置+長さ	85.9	83.1	56.9

定に対してモデルを個別に訓練させた場合とほぼ同じ性能が得られることを確認している。ただし、制御トークンを用いない生成（表1における「キーワード無し」）は上記のモデルとは独立してモデルの訓練と推論を行った。

キーワードの抽出 まず、ターゲットテキストから連続した1-3個の単語で構成されるフレーズをキーワード候補として抽出する。最初の単語がストップワードや頻出単語であるフレーズはキーワード候補から除外する。訓練時には、0-3個のフレーズをキーワード候補からエポック毎にランダムに選択し、キーワードとしてモデルに与える。推論時には、表1の実験では1-3個、表2と表4の実験では1個のキーワードをモデルに与える。

C 多様なテキストの生成

キーワードの位置を制御することで特定のキーワードから多様なテキストを生成できることを示す。多様なテキストを生成できれば、ユーザーは生成された複数のテキストの中から自分の意図するテキストを選択することが可能となる。キーワード無しの通常の生成とキーワードのみを用いた生成の場合、1つの入力からビームサーチもしくはtop-pサンプリングを用いて10種類のテキストを生成する。キーワード位置を用いる場合、1つのキーワードに対して10種類のキーワード位置をモデルに与えることで10種類のテキストを生成する。生成されたテキストに対してSelf-BLEU [24]指標を用いて生成テキストの多様性を評価する。

表6の結果から、複数のキーワード位置を用いて生成することで生成テキストの多様性が向上することが分かる。特に、要約タスクは一般的にビームサーチを用いてテキストを生成するため、複数テキストを生成した場合の多様性が非常に低くなるが、複数の位置を指定してテキストを生成することでこの多様性の低下が緩和される。表3に示した生成例においても、キーワード位置を制御することで特定のキーワードから様々なテキストが生成できることが確認できる。