# 指定要約長に応じたソフトな内容選択による要約長操作可能な End-to-End 要約

楢木 悠士 [1]　小林 哲則 [1]　林 良彦 [1]
[1*] 早稲田大学理工学術院 yuji.1277@akane.waseda.jp

## 概要

指定した長さに応じて適切な要約を生成する技術への要請が高まっている．本研究では，任意の指定要約長 $k$ が与えられたとき，原文書中のトークンの重要度を評価し，それに応じてソフトマスクをかけることで埋め込み表現を調整する機構 SumTop-$k$ を提案する．SumTop-$k$ は，要約長操作のための制約なしで，要約タスクによる End-to-End 学習によって要約長の制御を可能にした．評価実験の結果，提案手法は既存手法と同等以上の要約性能を達成した．本稿ではさらに，SumTop-$k$ 内で計算されるトークンごとのスコアが生成される要約に与える影響を分析し，これらを意図的に変化させることで生成要約の内容を操作できる可能性を示す．

## 1　はじめに

自動要約に関する研究は多く，要約技術の応用例も増えてきている．我々は，要約技術の実用化に向けて必要となる要約長の操作に注目する．長さ制御可能な要約は，所望の長さの簡潔で流暢な要約を生成することを目的としている．その技術は，各デバイス，デザイン，ユーザに応じた適切な長さの要約を提供するために，ユーザエクスペリエンスの観点から求められている．要約に含めるべき情報量や粒度は，指定要約長によって異なるため，このタスクは単に長さを制御するだけでは機能しない．したがって，要約長操作可能な要約システムでは，生成されるテキストの長さを制御するだけでなく，要約に含めるべき情報を適切に選択する必要がある．

本論文では，指定される任意の要約長 $k$ に対して原文書中のトークンの重要度を評価し，それに応じて埋め込み表現を調整するソフトマスク機構 SumTop-$k$ を提案する．ソフトマスク生成には微分可能な top-k 演算 [1] を用いることで，End-to-End 学習を可能にする．これは，内容選択機能を独立した抽出型要約モデルによって実現する既存研究 [2] における End-to-End 学習ができないという問題を解決する．

BART の Encoder に提案手法の SumTop-$k$ を組み込み，生成要約の品質と長さの操作性を CNN DailyMail と XSum データセットを用いた実験により評価した．その結果，提案手法は既存手法と同等以上の要約性能を達成しつつ，要約長の制御が可能であることを確認した．特に CNN DailyMail データセットにおいては BART や要約長操作可能な既存研究と比べて最も高い ROUGE スコアを達成し，XSum データセットでは，BART と同程度の ROUGE スコアを示した．さらに，SumTop-$k$ 内で計算されるトークンごとのスコアが生成される要約に与える影響を分析し，SumTop-$k$ 内のトークンスコアは生成要約に含まれるトークンと相関があることを示した．この結果を受け，SumTop-$k$ のスコアの順位換えによって，要約の内容を操作する可能性を検討した．

## 2　関連研究

要約長操作に関連する既存研究では，要約生成時に制御を行う手法が提案されてきた．菊池ら [3] は指定要約長までの残りの長さの埋め込みをモデルに入力することで，高瀬と岡崎 [4] は残りの長さを Transformer の位置埋め込み表現に追加で埋め込むことで要約長を操作した．Yu らは推論時に指定要約長の埋め込み表現を単語予測層の前に連結することで，要約長の制御を実現している [5]．しかし，指定要約長の情報をモデルに与えるだけでは，指定要約長に応じて適切に要約の内容を変化させることは困難である．すなわち，指定要約長に応じた内容選択が必要となる．

斉藤らは，抽出型要約モデルを用いて原文書から情報を選択し，原文と抽出されたトークンを抽象型要約モデルに入力する方式を提案した [2]．このア

プローチは，要約長の操作性と高い ROUGE スコアを達成したが，情報選択のための抽出型モデルが End-to-End に学習ができないため，性能上のボトルネックになっていると考えられる．一方，Liu らは内容選択に注力した手法を提案した [6]．このモデルでは，まず合成データによる事前学習によって内容選択機能を実現し，元のデータセット上で全体の要約タスクでの学習を行う．さらに，指定要約長付近で EOS トークンを高確率で生成するために，decoder における Attention スコアを指定要約長までの残りの長さに応じて調整している．この Attention スコアの調整方法は，生成要約長の操作に有効であるが，生成要約の品質の観点からは多分に強引であると考えられる．

我々は，原文書に対して，指定要約長 $k$ に応じたソフトマスクをかける機構 SumTop-$k$ を提案する．SumTop-$k$ は，内容選択に関する機能を End-to-End 学習が可能なモデルとしながら，Liu らのように直接的な確率調整を行わずに要約長制御を可能とする方法である．

## 3 提案手法: SumTop-$k$

本節では，指定要約長 $k$ に応じたソフトマスクを原文書にかける機構 SumTop-$k$ の構成について述べる．我々は，文書要約モデルとして Encoder-Decoder 型モデルの代表格である BART を採用する．SumTop-$k$ モジュールは，BART の Encoder の最終層を置き換えるものであり，SumTop-$k$ の入力にソフトマスクをかけた埋め込み表現を出力する．埋め込み表現の shape は，$[batch\_size, src\_len, embed\_dim]$ である．

Encoder 最終層における適用方法については複数の可能性があるが，本研究では図 1 に示す 2 つの適用方法について実験を行った．`Query` (図 1 の (a)) は，Self-Attention の `Query` の入力に SumTop-$k$ を適用したもので，原文書とマスクした表現の間の Source-Target Attention のような処理になる．`After_Attn` (図 1 の (b)) は，SumTop-$k$ を Self-Attention の後の埋め込み表現に適用することを意味し，Encoder の出力により直接的な影響を与える．

SumTop-$k$ モジュールのアーキテクチャを図 2 に示す．SumTop-$k$ は 3 段階の処理を持つ．初めに，Self-Attention と 1 層の線形層を用いて，$[batch\_size, src\_len, 1]$ の shape をしたトークンごとのスコアを計算する，次に，トークンごとのスコア

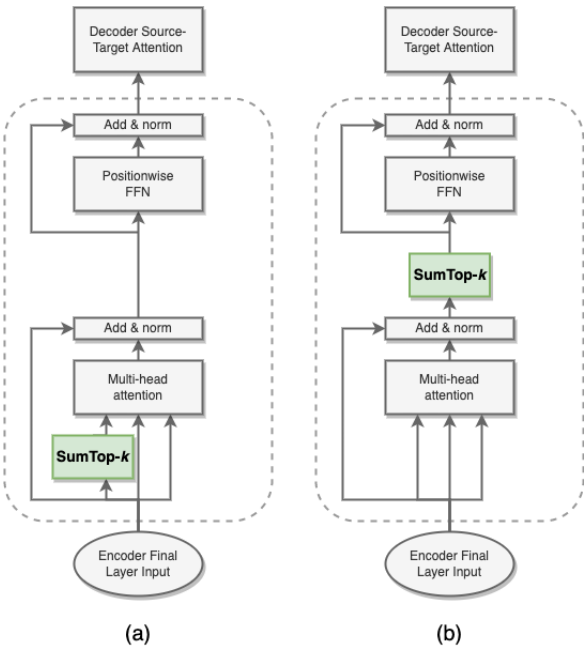


**図 1** SumTop-$k$ の適用方法．(a) と (b) はそれぞれ `Query` と `After_Attn` を表す．

を要約生成のための重要度だと捉え，それに応じたソフトマスクを微分可能な top-k 演算 [1] を用いて生成する．最後に SumTop-$k$ は元の入力表現にソフトマスクをかけた表現を出力する．SumTop-$k$ は完全に微分可能であり，モデルの学習時にどのようなトークンのスコアを高くすべきかなどの追加の制約は与えられることはない．よって，SumTop-$k$ がどのように原文書の表現を調整するかは，モデル全体の End-to-End の学習に依存する．

## 4 評価実験の設定

CNN DailyMail，XSum という要約研究における標準データセットを用い，要約品質と要約長制御に関する評価を行う．また，図 1 (a)，(b) に示した SumTop-$k$ の 2 つの配置について比較を行う．

CNN DailyMail データセットと XSum データセットでは，ビームサイズをそれぞれ 4 と 6 に設定した．また，ビームサーチ時の n-gram の繰り返し回数の最大値は 2 とした．手法自体の要約長の操作性を検証するため，提案手法で検証する際には，要約長の最大・最小の制約は与えない．

比較手法には，文書要約手法である BART および BART+attention head masking (AHM) [7]，要約長操作可能な要約手法である LPAS[2] および PtLAAM[6] を用いる．

要約の品質を測るために ROUGE スコアを，要約

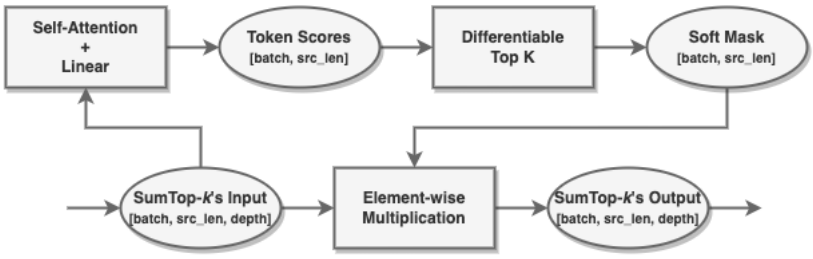


**図 2**　SumTop-$k$ の構造.

長の操作性を測るために生成要約の平均長と式 1 で定義される Variance を使用する．なお，$y_i$ は参照要約，$y'_i$ は生成要約，$n$ はテストデータ数，$len(\cdot)$ は入力に含まれるトークンの数を返す．

$$\mathrm{Var} = 0.001 * \frac{1}{n}\sum_{i=1}^{n}\left|len(y_i) - len(y'_i)\right|^2 \tag{1}$$

前述したように，SumTop-$k$ の処理で高いスコアを持つトークンには制約がないため，どのようなトークンが高いスコアを持つかを確認する必要がある．我々は，ソフトマスクの値が高いトークンは生成要約に現れやすいと仮説立てた．そこで，SumTop-$k$ の操作とモデルの最終出力の関係を分析するために，トークンの生成要約に対する貢献の度合いを表す，式 2 に示す指標 $Con$ を提案する．SumTop-$k$ のトークンスコアで上位指定要約長個のトークン ($T_{top-k}$) を SumTop-$k$ で重要だと判断されたトークンと考え，生成要約に含まれるトークン ($T_{gen}$) に対するトークンの精度を貢献の度合いとする．

$$\mathrm{Con} = \frac{count\left(set\left(T_{top\text{-}k}\right) \cap set\left(T_{gen}\right)\right)}{count\left(set\left(T_{gen}\right)\right)} \times 100 \tag{2}$$

ここで，$count(\cdot)$ は入力集合の要素数を返し，$set(\cdot)$ は入力配列の要素の集合を返す．

# 5　評価実験の結果

## 5.1　要約の品質と要約長の操作性

CNN DailyMail と XSum のデータセットに対して，提案手法と既存手法を評価した結果を表 1 に示す．本実験にける指定要約長は参照要約の長さとした．BART+AHM [7]，LPAS [2] と PtLAAM [6] の ROUGE スコアはそれぞれの論文から，LPAS と PtLAAM の Variance は Liu らが示した図 [6] を参照している．

まず，生成要約の品質について考察する．CNN DailyMail データセットにおいて，提案手法は既存手法と比較して高い ROUGE スコアを示した．特に，提案手法は `After_Attn` において，最も高い ROUGE スコアを示した．XSum データセットでは，提案手法はベースラインである BART と同程度の ROUGE スコアを示し，LPAS を上回った．XSum では CNN DailyMail と比較して ROUGE スコアが向上しなかったが，これは XSum データセットでは CNN DailyMail と比較して正解要約の抽象度が高いからだと考えられる．

次に，要約長の操作性について考察する．両データセットにおいて，提案手法はベースラインの BART よりも低い Variance を達成し，より良い操作性を持つことが確認できた，なお，Liu らの報告した BART の Variance と我々の測定した BART の Variance は大きく異なるため，本論文で比較するには公平でないと考え，提案手法と既存手法の Variance は比較しないこととした．

モデル内部の各トークンのスコアと生成要約の相関を測る貢献度では，提案手法のトークンスコアは BART の Attention スコアより高い．特に，CNN DailyMail の SumTop-$k$ `After_Attn` では，生成要約に含まれるトークンの 40%以上が SumTop-$k$ の上位 k 個のトークンスコアに含まれる．これらの結果から，提案手法は，要約に含めたいトークンにより他界スコアを与えていることが示唆される．

## 5.2　要約の内容操作の可能性

SumTop-$k$ を用いたモデルにおける要約の内容操作の可能性を検討する．提案手法による貢献度スコアが高いことは表 1 に示した通りである．これは，SumTop-$k$ におけるスコアの順位を意図的に変更することで，生成要約に現れるトークンが変化する可能性があることを示唆する．そこで，SumTop-$k$ は要約の内容を操作することができると仮説立てた．

**表 1** CNN DailyMail と XSum のデータセットに対する既存の 4 つの手法と 2 種類の SumTop-$k$ の結果.

| | CNN DailyMail | | | | | |
|---|---|---|---|---|---|---|
| | R-1 | R-2 | R-L | Mean Length | Variance | Con |
| BART | 43.87 | 21.12 | 40.64 | 83.5 | 1.201 | 14.43 |
| BART+AHM | 45.54 | 22.24 | 42.44 | - | - | - |
| LPAS | 43.23 | 20.46 | 40.00 | - | 0.253 | - |
| PtLAAM | 44.17 | 20.63 | 40.97 | - | **0.046** | - |
| SumTop-$k$ `Query` | 45.29 | 21.78 | 41.81 | 60.5 | 0.148 | 38.57 |
| SumTop-$k$ `After_Attn` | **46.00** | **22.33** | **42.61** | 67.3 | 0.231 | 43.07 |
| | XSum | | | | | |
| | R-1 | R-2 | R-L | Mean Length | Variance | Con |
| BART | 44.69 | 21.31 | 36.05 | 23.8 | 0.0452 | 16.03 |
| BART+AHM | 45.35 | 22.31 | 37.15 | - | - | - |
| LPAS | 43.23 | 20.46 | 40.00 | - | 0.296 | - |
| PtLAAM | **45.48** | **21.80** | **36.84** | - | 0.0162 | - |
| SumTop-$k$ `Query` | 44.50 | 20.82 | 35.40 | 23.6 | 0.0184 | 30.69 |
| SumTop-$k$ `After_Attn` | 44.17 | 20.49 | 35.14 | 22.6 | 0.0189 | 31.79 |

この仮説を検証するため，ROUGE スコアが低い (我々の分析では ROUGE-1 で 0.2 以下) テストデータに対して，意図的に順位を変更した場合の生成要約の変化を観察した．順位換えするトークンは，元の文書に出現したトークンのうち，正しい要約には出現するが生成要約には出現しないトークンとした．さらに，機能語ではなく内容語のみに変更を加えたいので，原文書に 3 回以上現れるトークンは順位換えの対象から除外した．CNN DailyMail データセットで SumTop-$k$ の `After_Attn` を適用した結果について分析を行った．SumTop-$k$ のスコアにおける順位を調整した後に，再度推論したときの生成要約がどのように変化するかを ROUGE スコアで比較した結果を表 2 に示す．具体的には指定要約長 $k$ が 10 のとき，順位の上げ幅は 10 倍の 100 とした．また，2 つの事例を表 3 に示す．

**表 2** 順位換えの前と後の ROUGE スコアの結果.

| | R-1 | R-2 | R-L |
|---|---|---|---|
| SumTop-$k$ `After_Attn` | 15.34 | 2.11 | 13.96 |
| SumTop-$k$ `After_Attn` Reranked | 19.53 | 5.76 | 18.15 |

**表 3** SumTop-$k$ のトークンスコアの順位換えによる要約の変化の事例.

| | |
|---|---|
| Before reranking | Solar Probe Plus will carry four experiments into the sun 's outer atmosphere . It will study the solar wind and energetic particles as they blast off the surface of the star . |
| Reranked | Solar Probe Plus will study the solar wind and energetic particles as they blast off the star . The launch window opens for 20 days starting July 31 , 2018 . |
| Reference | Temperatures outside the spacecraft will reach 2,500 degrees Fahrenheit . Launch window opens for 20 days starting on July 31 , 2018 . |

表 2 から，意図的な順位換えが ROUGE スコアを向上させたことが確認できる．つまり，この操作により生成要約の内容が参照要約の内容に近づいた．表 3 の例では，順位換えにより，生成要約に "launch window" に関する説明が含まれるようになった．

## 6 おわりに

我々は，原文書に対して，任意の指定要約長 $k$ に応じたトークンスコアの算出とそれに伴うソフトマスクをかける機構 SumTop-$k$ を提案した．提案手法は End-to-End での学習が可能であり，追加の制約なしに要約の長さを操作することが可能である．

評価実験の結果，提案手法は既存手法と同等以上の要約性能を達成しつつ，要約長の制御が可能であることを確認した．特に CNN DailyMail データセットにおいては BART や要約長操作可能な既存研究と比べて最も高い ROUGE スコアを達成し，XSum データセットでは，BART と同程度の ROUGE スコアを示した．また，SmTop-$k$ のトークンスコアで上位 k 個のトークンが，生成要約に含まれるトークンと相関を持つことを確認し，SumTop-$k$ を用いることが生成要約の内容に寄与していることを確認した．さらに，トークンスコアを意図的に順位換えすることで，低品質の要約を改善し，生成要約の内容を操作できることを示した．

このような要約制御の機能は，自動要約技術を実応用する上で重要と考えられる．今後は，原文書に現れないトークンが含まれるような抽象的な要約を生成できる長さ制御可能な要約手法を検討する．

# 参考文献


[1] Yujia Xie, Hanjun Dai, Minshuo Chen, Bo Dai, Tuo Zhao, Hongyuan Zha, Wei Wei, and Tomas Pfister. Differentiable top-k with optimal transport. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, **Advances in Neural Information Processing Systems**, Vol. 33, pp. 20520–20531. Curran Associates, Inc., 2020.

[2] 斉藤いつみ, 西田京介, 西田光甫, 大塚淳史, 浅野久子, 富田準二, 進藤裕之, 松本裕治. 出力長制御と重要箇所の特定を同時に行う生成型要約. 人工知能学会全国大会論文集, Vol. JSAI2020, pp. 3Rin481–3Rin481, 2020.

[3] Yuta Kikuchi, Graham Neubig, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura. Controlling output length in neural encoder-decoders. In **Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing**, pp. 1328–1338, Austin, Texas, November 2016. Association for Computational Linguistics.

[4] Sho Takase and Naoaki Okazaki. Positional encoding to control output sequence length. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 3999–4004, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[5] Zhongyi Yu, Zhenghao Wu, Hao Zheng, Zhe XuanYuan, Jefferson Fong, and Weifeng Su. LenAtten: An effective length controlling unit for text summarization. In **Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021**, pp. 363–370, Online, August 2021. Association for Computational Linguistics.

[6] Yizhu Liu, Qi Jia, and Kenny Zhu. Length control in abstractive summarization by pretraining information selection. In **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 6885–6895, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[7] Shuyang Cao and Lu Wang. Attention head masking for inference time content selection in abstractive summarization. In **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 5008–5016, Online, June 2021. Association for Computational Linguistics.

# A Case Study

| Source document (Bolded tokens are the tokens with top k scores of SumTop-k) |
|---|
| **Paul Walker** is hardly the first **actor** to die during a production. But **Walker**'s **death** in **November 2013** at the age of **40** after a **car crash** was especially eerie given his rise to fame in the " **Fast** and **Furious**" **film franchise**. The **release** of **" F urious 7**" on **Friday** offers the opportunity for fans to remember -- and possibly grieve again -- the man that so many have praised as one of the **nic**est guys in **Hollywood**. "He was a **person** of humility, integrity, and compassion," **military** veteran **Kyle** Upham said in an email to CNN. Walker **secretly paid** for the engagement ring Upham shopped for with his bride. "We didn't know him personally but this was apparent in the short time we spent with him. I know that we will never forget him and he will always be someone very special to us," said Upham. The actor was on **break** from **filming** "Furious 7" at the time of the fiery accident, which also claimed the life of the car's driver, Roger Rodas. Producers said early on that they would not **kill** off Walker's character, **Brian** O'Connor, a former cop turned road racer. Instead, the script was rewritten and **special** effects were used to finish scenes, with Walker's **brothers**, Cody and Caleb, serving as **body** doubles. There are scenes that will resonate with the audience -- including the **ending**, in which the filmmakers figured out a touching way to pay tribute to Walker while "retiring" his character. At the **premiere** Wednesday night in Hollywood, Walker's **co**-star and close friend **Vin** Diesel gave a tearful speech before the screening, saying "This movie is more than a movie." "You'll feel it when you see it," Diesel said. "There's something emotional that happens to you, where you walk out of this movie and you appreciate everyone you love because you just never know when the last day is you're gonna see them." There have been **multiple t**ributes to Walker **leading** up to the release. Diesel revealed in an interview with the "Today" show that he had named his newborn daughter after Walker. Social media has also been paying homage to the late actor. A week after Walker's death, about 5,000 people attended an outdoor **memorial** to him in Los Angeles. Most had never met him. Marcus Coleman told CNN he spent almost $1,000 to truck in a banner from Bakersfield for people to sign at the memorial. "It's like losing a friend or a really close family member ... even though he is an actor and we never really met face to face," Coleman said. "Sitting there, bringing his movies into your house or watching on TV, it's like getting to know somebody. It really, really hurts." Walker's younger **brother** Cody told People magazine that he was initially nervous about how "Furious 7" would turn out, but he is happy with the film. "It's bittersweet, but I think Paul would be proud," he said. CNN's Paul Vercammen contributed to this report. |
| Reference |
| "Furious 7" pays tribute to star Paul Walker, who died during filming . Vin Diesel: "This movie is more than a movie" "Furious 7" opens Friday . |
| BART |
| Paul Walker's death in November 2013 at the age of 40 after a car crash was especially eerie given his rise to fame in the "Fast and Furious" film franchise . The release of "Furious 7" on Friday offers the opportunity for fans to remember -- and possibly grieve again . There have been multiple tributes to Walker leading up to the release . |
| BART + SumTop-k desired length: 37 (gold length) |
| Paul Walker's death in November 2013 at the age of 40 after a car crash was especially eerie given his rise to fame in the "Fast and Furious" film franchise . The actor was on break from filming "Furious 7" at the time . |
| BART + SumTop-k desired length: 10 |
| Walker died in a car crash in November 2013 . |
| BART + SumTop-k desired length: 30 |
| Paul Walker's death in November 2013 at the age of 40 after a car crash was especially eerie given his rise to fame in the "Fast and Furious" film franchise . |
| BART + SumTop-k desired length: 50 |
| Paul Walker's death in November 2013 at the age of 40 after a car crash was especially eerie given his rise to fame in the "Fast and Furious" film franchise . The release of "Furious 7" on Friday offers the opportunity for fans to remember -- and possibly grieve again . |
| BART + SumTop-k desired length: 70 |
| Paul Walker's death in November 2013 at the age of 40 after a car crash was especially eerie given his rise to fame in the "Fast and Furious" film franchise . Producers said early on that they would not kill off Walker's character, Brian O'Connor, a former cop turned road racer . The actor was on break from filming "Furious 7" at the time of the fiery accident . |

**図 3** 原文書と参照要約の例と，BART と 5 種類の異なる指定要約長での提案手法による生成要約.

図 3 には，CNN DailyMail データセットの原文書とその参照要約と，BART と我々の提案手法である BART + SumTop-$k$ `After_Attn` により，5 種類の異なる指定要約長の生成要約を示した．原文書の太字のトークンは SumTop-$k$ のトークンスコアが上位 k 個のトークンである．

BART と比較して，提案手法 SumTop-$k$ で指定要約長を参照要約の長さとすると，要約が短くなり参照要約の長さに近づくことがわかった．指定要約長が 10 と 30 の生成要約では，ポール・ウォーカーの事故について記述されているが，指定要約長が 50 の要約では，シリーズの新作について，指定要約長が 70 の要約では，製作者の発言について記述されている．このように，希望する長さによって，要約に含めるべき内容が異なることが確認された．SumTop-$k$ のスコアが上位 k 個のトークン（太字部分）について前述の貢献度の結果からわかるように，SumTop-$k$ のスコアが上位 k 個のトークンも生成要約に含まれる．