

重要文指定を行う Transformer による文書要約

鈴木 佑真[†] 山本 幹雄[‡]

筑波大学 理工情報生命学術院 システム情報工学研究群[†] 筑波大学 情報システム系[‡]

1 はじめに

近年ニュースや SNS などの、Web 上のテキストデータはますます増加しており、それらのコンテンツを全て読んで必要な情報を得るには多大な時間を要する。そこで効率的に必要な情報を得るためにはコンテンツの要点をまとめた要約文が有用であり、それに伴い文書の自動要約技術が様々な観点から研究されている[1]。

本研究では、近年自然言語処理の分野において高い成果を挙げている Transformer[2]と呼ばれるニューラルネットワークモデルに対して、文書中の重要文を明示的に指定する機構を追加することで、要約の高精度化を目指す。本稿において、Transformer に文書中の重要文を指定するための具体的な機構と、提案機構を組み込んだモデルの評価実験について述べる。

2 Gap-Sentences Generation

Gap-Sentences Generation(GSG)とは文書要約に特化した Transformer の事前学習方法である[1]。GSG では、 n 個の文で構成されている文書 $D = \{x_i\}_n$ に対して、式(1)の値が高い順にマスクして Transformer に入力を行う。

$$s_i = ROUGE\{x_i, D - \{x_i\}\} \quad (1)$$

ROUGE とはテキスト要約の評価指標であり[3]、0.0 から 1.0 の数値を用いて人間が作成した要約とシステムが生成した要約の一致度を測る。このときこの数値が高いほど2つの要約は一致度が高いといえる。式(1)では、文書全体との一致度が高い文 x_i に対応する s_i の値が高くなる。

ここで s_i の値が高いとき、対応する文 x_i は文書 D の内容をよく表している重要な文であると考えることができる。本研究ではこの方法で要約対象の文書中から重要な文を計算し、それを Transformer の学習時または推論時に明示的に指定することで、要約の精度の改善を図る。

3 提案手法

本研究では、入力文書中の各文が重要である

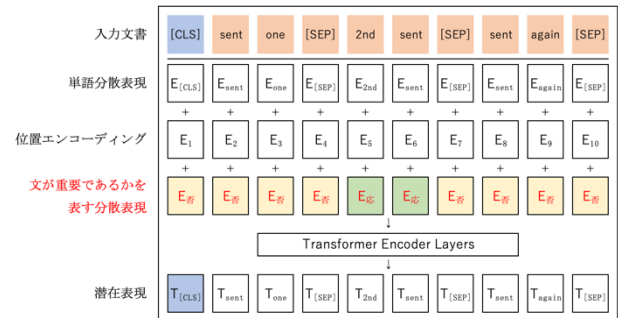


図1 文が重要か否かを表す分散表現の付加

か否かを Transformer に明示的に指定することで要約の高精度化を図る。具体的には図1に示すように、Transformer の Encoder への入力である各単語の分散表現に対し、入力文書中の各文が重要であるか否かを表す分散表現を、位置エンコーディング[2]と同じタイミングで加算する。入力文書中の各文が重要であるか否かを決定するには式(1)を用い、この値が高い文を重要な文とする。

図1では、3つの文が入力されていて、2つ目の文が重要である場合の例である。2つ目の文に対応する単語の分散表現に対しては、その文が重要であるということを示す分散表現が付加されており、それ以外の単語の分散表現に対しては、その文が重要ではないということを示す分散表現が付加されている。このように文が重要であるかを表す分散表現は2種類のパターンをもち、そのどちらかが各単語の分散表現に付加される。これらの各文が重要かどうかを表す分散表現は、Transformer の学習と同時に学習させる。本稿ではこのモデルを提案モデル1と呼ぶ。

また同時に、図2に示すように Decoder の入力には文が重要であることを表す分散表現を付加することを考える。これにより、Encoder で重要と指定した文と Decoder の入力の関連度がより高くなり、重要な文に対して注意が向けられるようになることが期待できる。本稿ではこのモデルを提案モデル2と呼ぶ。

4 評価実験

本研究では従来の Transformer (従来モデル) と提案モデル1、提案モデル2を同条件下で学習さ

[†]Yuma SUZUKI, Graduate School of Science and Technology, Degree Programs in Systems and Information Engineering, University of Tsukuba

[‡]Mikio YAMAMOTO, Faculty of Engineering, Information and Systems, University of Tsukuba

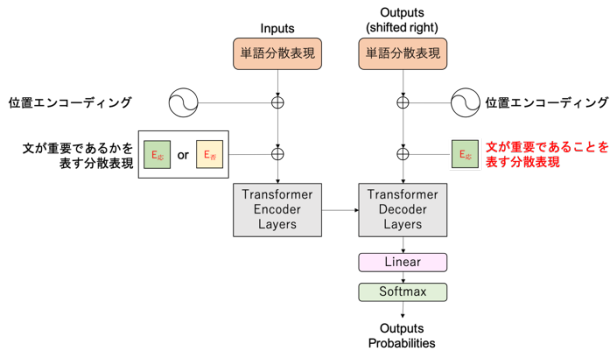


図2 Decoderの入力への分散表現の付加

せ、それぞれのモデルが出力した要約と、人手要約の ROUGE スコアを計算し、一致度を測ることでモデルの良し悪しを比較する。このとき ROUGE の中でも ROUGE-1, ROUGE-L の 2 つの指標を用いて評価を行った。また提案モデルについては式 (1) を用いて重要な文を計算する必要があるが、これには ROUGE-L を使用して値を求めた。さらに提案モデルに指定する重要な文の数を 0 から 8 の場合と全ての文を重要と指定する場合のそれぞれについて学習・評価を行った。

4.1 評価データセット

本実験では livedoor ニュース¹からニュース記事とそれに対応する要約を 134,000 ペア収集し、日本語要約データセットを構築した。このデータセットを 130,000 件の学習データ, 2,000 件の検証データ, 2,000 件のテストデータに分割し、学習データ/検証データを用いてモデルの学習を行った上で、テストデータを用いて実際に要約を生成し、ROUGE スコアを測定する。

4.2 実験結果

従来モデル, 提案モデル 1, 提案モデル 2 のそれぞれについてモデルの学習を行い、テストデータを用いて要約を生成した。人手の要約と生成した要約の ROUGE-1, ROUGE-L スコアを図 3, 図 4 に示す。ここで N は 1 つの入力データに対して重要と指定した文の数である。また全ての文を重要と指定した場合の結果を $N=\infty$ として記している。

提案モデル 1 と提案モデル 2 の ROUGE スコアは共に従来モデルと比べておおよそ高い水準となったが、提案モデル 1 に比べて提案モデル 2 のスコアの向上度が顕著であった。Decoder の入力に文が重要であることを表す分散表現を付加することで、Encoder で重要と指定した文に対してより注意が向けられたと考えられる。

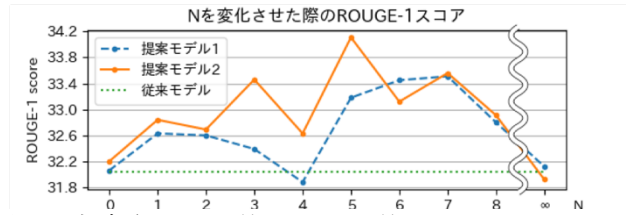


図3 生成された要約と人手要約の ROUGE-1 スコア

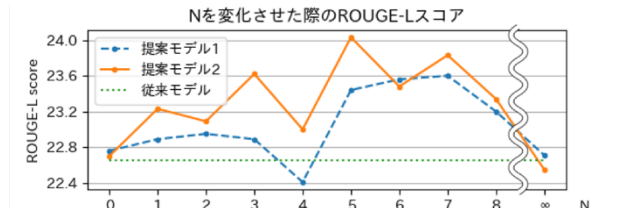


図4 生成された要約と人手要約の ROUGE-L スコア

また提案モデル 2 について、 $N=0, \infty$ を除いた 8 つの場合においては、有意水準 0.05 で従来モデルとの間に有意な差が見られた。特に $N=5$ の場合のスコアは最も高く、従来モデルのスコアと比べて ROUGE-L で 1.38 ポイント上昇した。 $N=0$ もしくは $N=\infty$ の場合は従来モデルと同等の結果となったが、 $N=1$ の場合は従来モデルと比べて ROUGE-L で 0.58 ポイント上昇した。1 文でも重要と指定する文が存在することにより、2 種類の分散表現それぞれに別の意味が付与されるように学習が行われたと考えられる。

5 おわりに

本稿では Transformer を用いた文書要約においてその精度を向上させるために、Transformer に対し入力文書中の重要な文を明示的に指定する手法を提案し、その有効性を示した。今後は重要な文の決定方法として他の手法を用いた場合の結果の変化や、Transformer の事前学習モデルに対する本手法の適用可能性について検討を行う。

参考文献

- [1] J. Zhang, M. Saleh and P. J. Liu : PEGASUS: Pretraining with Extracted Gap-sentences for Abstractive Summarization, ICML 2020, pp.11328-11339, 2020.
- [2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaise and I. Polosukhin : Attention is all you need, NIPS 2017, pp.5998-6008, 2017.
- [3] C. Lin and E. Hovy : Automatic Evaluation of Summaries Using N-gram Co-Occurrence Statistics, HLT-NAACL 2003, pp.71-78, 2003.

¹ <http://news.livedoor.com/>