

日本語ニュース記事要約支援に向けた ドメイン特化事前学習済みモデルの構築と活用

石原祥太郎¹ 村田栄樹² 高橋寛武³ 中間康文³
¹ 株式会社日本経済新聞社 ² 早稲田大学 ³ 独立研究者
shotaro.ishihara@nex.nikkei.com

掲載号の情報

31 巻 4 号 pp. 1717-1745.

doi: <https://doi.org/10.5715/jnlp.31.1717>

概要

本応用システム論文では、ドメイン特化事前学習済みモデルの産業応用として、日本語金融ニュースメディア「日経電子版」のニュース記事を自動要約する編集支援システム¹⁾の開発事例を報告した。記事の自動要約はニュースメディアにおける編集者の労働負荷の軽減に寄与する。しかしニュースメディアには独自の表記規定が数多く存在するため、汎用的なモデルの出力では不十分な場合がある。我々はドメイン特化事前学習済みモデルを構築し利用することで、より用途に適したシステムが構築できると実証した。

具体的にはまず、実際の編集現場の要請に基づく実社会のシステム要件を以下の4点に整理した。次にそれぞれの要件に対して、既存技術を組み合わせて開発した編集支援システムの有用性を検証した。

1. 特有の文体の再現 編集作業の負荷は、生成された要約を修正にかかる手間に依存する。本研究では、特有の文体を学習させる目的で、日経電子版のニュース記事を用いて T5 をフルスクラッチで事前学習した (t5-base-japanese-nikkei)。この T5 を、日経電子版のニュース記事の本文と 3 行要約の対でファインチューニングした結果、一般的なモデルを上回る自動要約の性能が確認できた (表 1)。見出し生成でも同様の傾向が観測された。

2. 性能と運用コスト 産業応用では、性能と運用コストの両面が重要視される。ドメイン特化事前学習済みモデルは特定のドメインに絞るため、一般的なモデルに比べて事前学習コーパスが

1) 本稿では技術検証の過程を報告しており、実際に利用されている編集支援システムとは異なる。

表 1 モデルごとの 3 行要約生成の性能比較。太字は最良の値を示す。

| ファインチューニングに用いた 事前学習済みモデルと、外部 API | ROUGE | | |
|-------------------------------------|--------------|--------------|--------------|
| | 1 | 2 | L |
| t5-base-japanese-nikkei | 0.472 | 0.286 | 0.371 |
| sonoisa/t5-base-japanese | 0.424 | 0.252 | 0.338 |
| megagonlabs/t5-base-japanese-web | 0.410 | 0.235 | 0.315 |
| gpt-3.5-turbo (ゼロショット) | 0.387 | 0.181 | 0.262 |
| gpt-3.5-turbo (3 ショット) | 0.396 | 0.185 | 0.273 |
| gpt-3.5-turbo (6 ショット) | 0.394 | 0.184 | 0.272 |

小規模になる。ここで t5-base-japanese-nikkei と megagonlabs/t5-base-japanese-web の事前学習コーパスはそれぞれ 3.37 GB と 784 GB であった。より小規模の事前学習コーパスで優れた性能が得られており、前処理の簡略化やストレージの運用コストといった実用上の価値が確認された。

3. 忠実性 ニュースは入力に忠実な要約が要求されるため、幻覚の問題は注意深く監視する必要がある。本研究では、発生し得る幻覚の特徴を明らかにするために、3 行要約生成にファインチューニングした t5-base-japanese-nikkei の出力を定量的・定性的に分析した。具体的には、用意した 97 本の記事に対してそれぞれ 3 つの異なる要約を生成し、人手で幻覚の有無や種類をアノテーションした。幻覚を含む場合、含まない場合と比較して生成時の平均情報量 (エントロピー) が有意に大きく、学習時と推論時のデータの性質の違いが幻覚の発生に繋がっている可能性も示唆された。

4. システム全体の有用性 システム全体ではクリック率予測や後編集の機能も備えている。クリック率予測に向けては、日経電子版のニュース記事で事前学習した BERT を構築し、実際のアクセスログを用いてファインチューニングし性能を検証した。自動生成された要約が幻覚を含み得るため、編集者が出力を確認し、必要に応じて後編集する過程を想定している。このような後編集の需要も含めて、システム全体の有用性について、社内展開した際の反響を踏まえて議論した。