

press release

朝日新聞

2022年4月27日

朝日新聞社 〒104-8011 東京都中央区築地 5-3-2

〈報道関係のみなさま〉

長い文章から要約文や見出しを自動で生成

自動要約生成 API「TSUNA®」のモデルをバージョンアップ

含めたい語句の指定が可能に 生成速度も向上

株式会社朝日新聞社(代表取締役社長:中村史郎)は、自然言語処理研究^(注1)の成果の一つである、自動要約生成 API^(注2)「TSUNA®」(https://cl.asahi.com/api_data/headlinegeneration.html)で利用している機械学習のモデルをバージョンアップし、公開しました。

(注1)「自然言語」は、人間が使う言葉の総称で、自然言語処理は人工知能(AI)を支える研究分野の一つです。

(注2)APIは、Application Programming Interface の略です。

「TSUNA®」は、長い文章を入力すると要約文や見出しを自動で生成します。文字数や割合を指定する事ができるので、用途に応じた文章を生成できます。この技術は過去30年分の朝日新聞記事のデータを活用し、学習データとしての処理効率を上げるためのフィルタリングなどをした上で、ディープラーニングにより可能となりました。この技術の一部は、特許を出願中です。



今回のバージョンアップでは、生成された要約文中に必ず含めたい語句の指定が可能になりました。この機能を実現する技術は論文化され、自然言語処理分野の最難関国際学会の1つである「EMNLP」に採択されています^(注3)。また、前バージョンよりも学習データの量を増やし、生成の速度も向上しています。

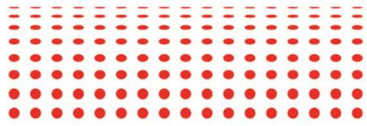
(注3) K. Yamada et al. Transformer-based Lexically Constrained Headline Generation. In Proceedings of the 2021 Conference on EMNLP, November 2021.

>

contact

〈お問い合わせ〉

朝日新聞社 メディア研究開発センター
mrad-contact@asahi.com



press release

朝日新聞

2022年4月27日

朝日新聞社 〒104-8011 東京都中央区築地 5-3-2

必ず含める語句を指定した場合の要約文の生成例:

(本文)

20日の外国為替市場で円相場が一時、約20年ぶりに1ドル=129円台まで下落した。19日の米ニューヨーク市場で128円90銭台まで下落していたが、その後のシドニー市場で129円台をつけた。1日で2円ほどと、急速に円安が進んだ。東京市場では128円台前半で推移している。

日本銀行は20日午前、長期金利の上昇を抑えるため、特定の利回りを指定して国債を無制限に買い入れる「指し値オペ」をすると市場に通知した。さらなる円安につながる可能性が指摘されていたが、市場は今のところ大きく反応していない。

...

(要約文の生成例)

指定語句無し ▶ 円、20年ぶり129円台 東京は128円台前半
「日本銀行」を指定 ▶ 円安、一時129円台 日本銀行、指し値オペ通知

「TSUNA®」は2019年に公開され、朝日新聞社内では広く業務に活用されている他、社外での利用も広がっています。また「TSUNA®」とともに、より長い文章を要約する「長文要約生成 API」(https://cl.asahi.com/api_data/longsum.html)の機械学習モデルも、学習データの量を増やしバージョンアップしています。

■朝日新聞社 メディア研究開発センターとは (<https://cl.asahi.com>)

朝日新聞社では、「メディア研究開発センター」が2021年4月に発足しました。人工知能を始めとする先端メディア技術と、新聞社ならではの豊富なテキストや写真、音声などの資源を活用し、社内外の問題解決を目指すとともに、自然言語処理や画像処理をはじめとした先端技術の研究・開発を進めています。(お問い合わせ先:mrad-contact@asahi.com)

■自動要約生成 API 「TSUNA®」デモサイト

https://cl.asahi.com/api_data/headlinegeneration.html

■「長文要約生成 API」デモサイト

https://cl.asahi.com/api_data/longsum.html

*利用にはお問い合わせフォームよりAPIキーの取得が必要です。

>

contact

〈お問い合わせ〉

朝日新聞社 メディア研究開発センター
mrad-contact@asahi.com