

要約作業を効率化する多機能な ニュース記事自動要約AIシステム

ドコモ北京研究所

ぐお
郭
り
李しゅーほん
塚
宏
あんしん
安
新なかむら
中
村
いっせい
一
成

サービスイノベーション部

ふじもと
藤
本ひろし
拓

現在、多くのメディアのニュースでは、記事の要約を表示することにより読者に内容を分かりやすく伝えている。しかし、人手による要約作業は分量が多い上に専門的なスキルを必要とし、人材確保の困難さの観点で課題となっている。そこでドコモは、要約作業を効率化できるよう、ユーザの意向に沿った要約を自動で作成する多機能な自動要約AIシステムを開発した。これにより、要約作業時間を短縮し、人手不足を解消することが可能となる。

1. まえがき

現在、多くのメディアのニュースでは、記事原稿の主要な内容を端的にまとめた要約を表示することにより、読者に内容を分かりやすく伝えている。しかし要約には、配信するメディアによって字数制限が存在することが多く、人手で必要な情報量を適切な長さに要約する作業は、分量が多い上に専門的なスキルを必要とする。そのため、スタッフの育成に一定期間の研修が必要な場合もあり、人材確保の観点で課題となっている。

このような課題を解決するために、近年、要約作業を効率化できるAIを用いたさまざまな自動要約システムがリリースされている。これらのシステムの多くは、原文から主要な文を抽出する抽出式要約や、原文から一部の単語やフレーズを削除、または追加することで新たな文を作成する生成式要約を提供している。抽出式要約と生成式要約の例を図1に示す。

生成式要約については、生成する要約の長さに合わせて一部の単語やフレーズを削除、または生成するため、既存技術では文法に誤りが生じることや、

©2022 NTT DOCOMO, INC.

本誌掲載記事の無断転載を禁じます。

本誌に掲載されている社名、製品およびソフトウェア、サービスなどの名称は、各社の商標または登録商標。

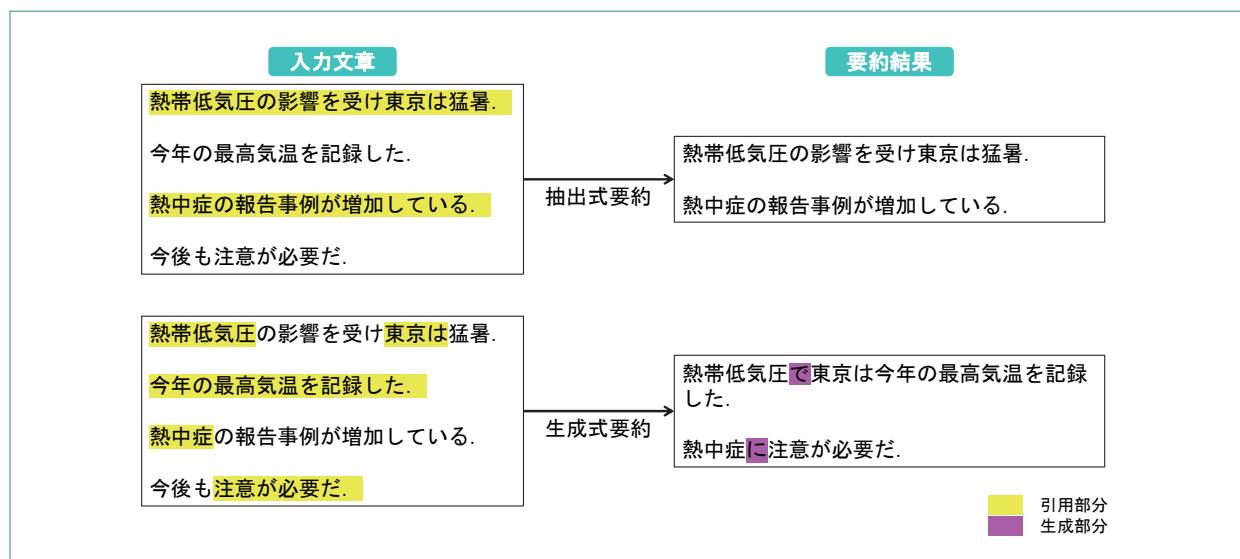


図1 抽出式要約と生成式要約の例

文法は正しい一方で指定の長さから逸脱してしまうことが課題である。また、既存の要約システムの多くは、ユーザの含めたいキーワードなどを考慮する機能や、要約内容が入力文章のどの位置に由来するかを特定し確認する機能がないため、実際の運用時には、要約内容に対して事前に制約を加えることができない点や、要約後に文法などに誤りがある場合に効率的に人手で修正することができない点も課題である。

そこでドコモは、上記の課題を解決できるユーザの意向に沿った要約を作成する多機能な自動要約AIシステムを開発した。本システムは特に3つの特長をもつ。1つ目は、要約の文字数制御機能についてユーザが指定した文字数の70~100%の長さの要約を作成できるようチューニングし、既存技術と比較して文字数制御の品質と要約の文法などの品質を大幅に向上させた点である。2つ目は、ユーザが目的の要約を作成しやすいように、要約に含める、もしくは含めないキーワードが設定できるヒント機能や、タイトルの内容を含む要約が生成できるタイトル機能を具備している点である。3つ目は、要約の

結果が正しいか確認、または修正といったユーザの作業を効率化できるよう、原文の抽出箇所と要約後の文章の生成箇所を可視化することができる位置特定機能を具備している点である。

本システムを利用することで、文字数や内容の観点でユーザの求める最適な要約を自動で生成し、生成後に人手による修正が必要な場合でも効率的に修正作業ができるため、人手による要約作成や従来の要約システムを利用した場合と比較して要約作業に要する時間を短縮し、人手不足を解消することが可能となる。

本稿では、ドコモの自動要約AIシステムの各機能の内容、生成式要約の技術的な性能改善方法、要約性能について解説する。

2. ドコモの自動要約AIシステム

2.1 システム全体像

ドコモの自動要約AIシステムには、深層学習^{*1}を用いた抽出式要約と生成式要約の2つのシステムがある（図2）。各システムに実装されている機能

*1 深層学習：多層のニューラルネットワークを用いた機械学習の一種。

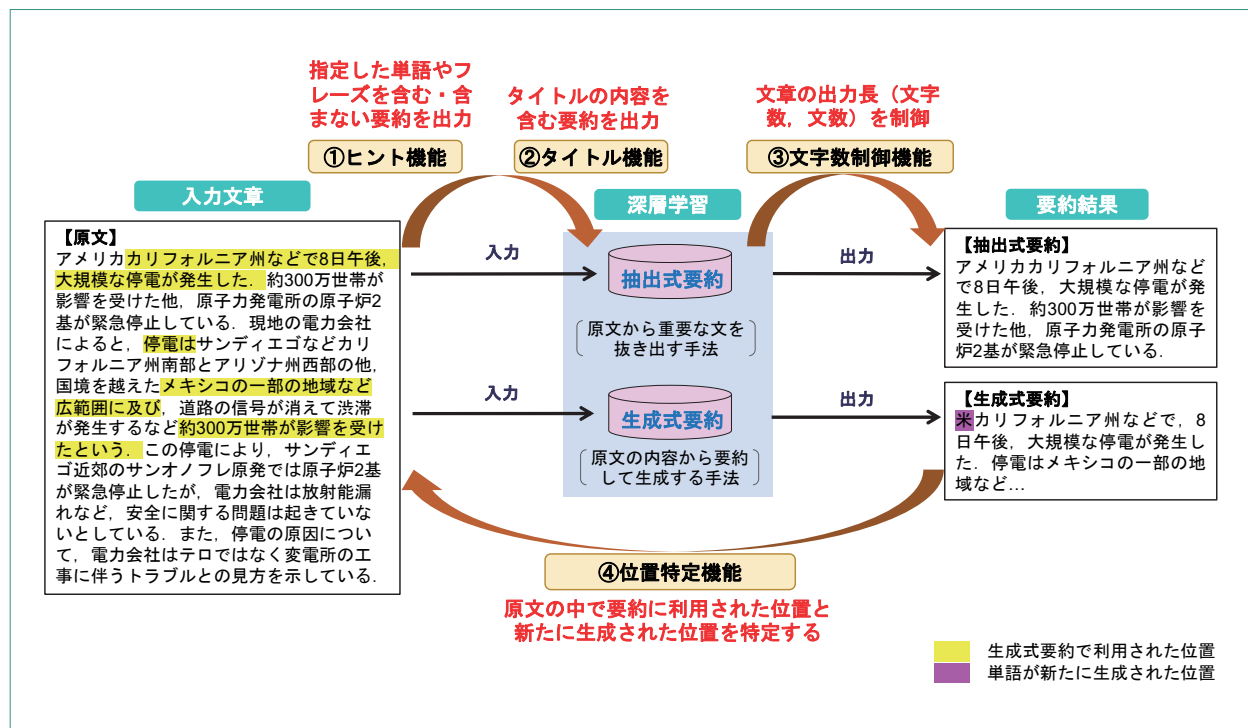


図2 ドコモの自動要約AIシステムの全体像

を以下に示す。

- ①ヒントとなるキーワードやフレーズを指定して要約内容に制約を加えるヒント機能
- ②タイトルの内容を含む要約を生成するタイトル機能
- ③要約の文字数や文の数を制御する文字数制御機能
- ④原文上の引用位置や新たに生成された文字の位置を可視化する位置特定機能

2.2 ヒント機能

ヒント機能は、ユーザが要約に含めたい、もしくは含めたくないキーワードやフレーズを指定することで、要約内容に制約を加えることができる機能である。ニュース記事において、登場人物、地域名、企業名、などさまざまな情報の中からどの情報を要約に含めたいかはユーザごとに異なる可能性があるため、各ユーザに対して最適な要約を生成できる機

能を開発した。

本システムでは、複数のキーワードやフレーズを入力することが可能である。図3は要約に含めたいキーワードをユーザが指定することで、ヒントが無い場合には要約に含まれない内容に含まれるように変化させた例である。図4は要約に含めたくないキーワード（図中ではマイナスヒントと記載）をユーザが指定することで、ヒントがない場合に要約に含まれていた内容が、含まれないように変化した例である。このようにユーザが入力するヒントによって、ユーザの意向により近い要約を出力することが可能である。

2.3 タイトル機能

タイトル機能は、ユーザが原文のタイトルを提供することにより、要約の質を向上させることができる機能である。通常、深層学習による要約の場合、

原文	指定文字数	ヒント指定なし 要約結果
<p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。約300万世帯が影響を受けた他、原子力発電所の原子炉2基が緊急停止している。現地の電力会社によると、停電はサンディエゴなどカリフォルニア州南部とアリゾナ州西部の他、国境を越えたメキシコの一部の地域など広範囲に及び、道路の信号が消えて渋滞が発生するなど約300万世帯が影響を受けたという。この停電により、サンディエゴ近郊のサンオノフレ原発では原子炉2基が緊急停止したが、電力会社は放射能漏れなど、安全に関する問題は起きていないとしている。また、停電の原因について、電力会社はテロではなく変電所の工事に伴うトラブルとの見方を示している。(294文字)</p>	120	<p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。停電はメキシコの一部の地域など広範囲に及び、約300万世帯が影響を受けたという。また、停電の原因について、電力会社はテロではなく変電所の工事に伴うトラブルとの見方を示している。(118文字)</p>
		<p>ヒント指定 “緊急停止” 要約結果</p> <p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。この停電により、サンディエゴ近郊のサンオノフレ原発では原子炉2基が緊急停止したが、電力会社は放射能漏れなど、安全に関する問題は起きていないとしている。(107文字)</p>

■ ヒント

図3 ヒント機能を利用したヒントを含む生成式要約結果の例

原文	指定文字数	マイナスヒント指定なし 要約結果
<p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。約300万世帯が影響を受けた他、原子力発電所の原子炉2基が緊急停止している。現地の電力会社によると、停電はサンディエゴなどカリフォルニア州南部とアリゾナ州西部の他、国境を越えたメキシコの一部の地域など広範囲に及び、道路の信号が消えて渋滞が発生するなど約300万世帯が影響を受けたという。この停電により、サンディエゴ近郊のサンオノフレ原発では原子炉2基が緊急停止したが、電力会社は放射能漏れなど、安全に関する問題は起きていないとしている。また、停電の原因について、電力会社はテロではなく変電所の工事に伴うトラブルとの見方を示している。(294文字)</p>	100	<p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。停電はメキシコの一部の地域など広範囲に及び、道路の信号が消えて渋滞が発生するなど約300万世帯が影響を受けたという。(90文字)</p>
		<p>マイナスヒント指定 “渋滞” 要約結果</p> <p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。この停電により、サンディエゴ近郊のサンオノフレ原発では原子炉2基が緊急停止したが、電力会社は、安全に関する問題は起きていないとしている。(100文字)</p>

■ マイナスヒント

図4 ヒント機能を利用したマイナスヒントを含まない生成式要約結果の例

要約の品質は学習データの量と質に依存し、学習データに含まれないような内容の文章が入力された場合に、要約の品質が低下することがある。そこで、タイトルに含まれるキーワードは原文の中でも特に重要な内容を抽出したものであると考え、タイトルをヒントとして与え、要約の質を高めるタイトル機能を開発した。図5に示すように、タイトルを指定することで、タイトル内のキーワードが含まれる要

約を生成することが可能である。

2.4 文字数制御機能

文字数制御機能は、ユーザが要約の文字数を指定できる機能である。ニュースサイトやソーシャルメディアには、配信メディアの要約表示枠のサイズの制約から要約文字数の最大値が厳しく制限されることが多いため、本システムではユーザが設定した文

字数の70～100%の範囲で要約を出力するようにチューニングをしている。図6に示すように同じ原文に対して異なる文字数を設定すると、それに合った文字数の要約をそれぞれ得ることができる。この機能を用いることで、さまざまな長さの要約を生成

することが可能である。なお、文字数だけでなく文の数を指定することも可能である。

2.5 位置特定機能

位置特定機能は、要約文言の原文上の引用位置や

入 力	指定文字数	タイトル指定なし 要約結果
<p>【タイトル】 米国などで大規模停電、原子炉2基緊急停止</p> <p>【原文】 アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。約300万世帯が影響を受けた他、原子力発電所の原子炉2基が緊急停止している。現地の電力会社によると、停電はサンディエゴなどカリフォルニア州南部とアリゾナ州西部の他、国境を越えたメキシコの一部の地域など広範囲に及び、道路の信号が消えて渋滞が発生するなど約300万世帯が影響を受けたという。この停電により、サンディエゴ近郊のサンオノフレ原発では原子炉2基が緊急停止したが、電力会社は放射能漏れなど、安全に関する問題は起きていないとしている。また、停電の原因について、電力会社はテロではなく変電所の工事に伴うトラブルとの見方を示している。(294文字)</p>	100	<p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。停電はメキシコの一部の地域など広範囲に及び、道路の信号が消えて渋滞が発生するなど約300万世帯が影響を受けたという。(90文字)</p>
		タイトル指定あり 要約結果
		<p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。この停電により、サンディエゴ近郊のサンオノフレ原発では原子炉2基が緊急停止したが、電力会社は、安全に関する問題は起きていないとしている。(100文字)</p>

■ ■ ■ タイトル関連キーワード

図5 タイトル機能を利用した生成式要約結果の例

原 文	指定文字数	要約結果
<p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。約300万世帯が影響を受けた他、原子力発電所の原子炉2基が緊急停止している。現地の電力会社によると、停電はサンディエゴなどカリフォルニア州南部とアリゾナ州西部の他、国境を越えたメキシコの一部の地域など広範囲に及び、道路の信号が消えて渋滞が発生するなど約300万世帯が影響を受けたという。この停電により、サンディエゴ近郊のサンオノフレ原発では原子炉2基が緊急停止したが、電力会社は放射能漏れなど、安全に関する問題は起きていないとしている。また、停電の原因について、電力会社はテロではなく変電所の工事に伴うトラブルとの見方を示している。(294文字)</p>	80	<p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。停電はメキシコの一部の地域など広範囲に及び、約300万世帯が影響を受けたという。(70文字)</p>
<p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。約300万世帯が影響を受けた他、原子力発電所の原子炉2基が緊急停止している。現地の電力会社によると、停電はサンディエゴなどカリフォルニア州南部とアリゾナ州西部の他、国境を越えたメキシコの一部の地域など広範囲に及び、道路の信号が消えて渋滞が発生するなど約300万世帯が影響を受けたという。この停電により、サンディエゴ近郊のサンオノフレ原発では原子炉2基が緊急停止したが、電力会社は放射能漏れなど、安全に関する問題は起きていないとしている。また、停電の原因について、電力会社はテロではなく変電所の工事に伴うトラブルとの見方を示している。(294文字)</p>	160	<p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。停電はメキシコの一部の地域など広範囲に及び、道路の信号が消えて渋滞が発生するなど約300万世帯が影響を受けたという。また、停電の原因について、電力会社はテロではなく変電所の工事に伴うトラブルとの見方を示している。(136文字)</p>

■ 引用部分

図6 文字数制御機能を利用した生成式要約結果の例

新たに生成された文字の位置を可視化する機能である。既存の自動要約AIシステムでは、出力する要約が原文のどこに位置するか表示されないため、ユーザが原文と要約結果を比較して、要約が原文の中の重要な内容を含み、文法的に正しいかどうかを確認する作業に時間を要する。そこで、ユーザが効率的に原文と要約結果を比較できるように、原文の中で要約として引用された文言の位置と、要約の中で新たに生成された文言の位置を特定できる位置特定機能を開発した。図7に示すように、「原子炉2基が緊急停止」という同じフレーズが原文に複数ある場合や、単語が新たに生成された場合にも正しく参照箇所をマッピングすることが可能である。

3. 生成式要約の品質改善手法

3.1 概要

通常、深層学習を用いた要約モデルの学習では、原文と人手で作成した正しい要約のペアのデータを用いて学習を行うが、人手で要約を作成するのは時間的、また金銭的なコストがかかるため、大量のデータが得られず性能の向上が図れないことが課題である。

そこで、ドコモの自動要約AIシステムでは、大

量のデータを用意する代わりに、原文と要約のペアデータから機械的に生成された誤りのある要約や、正解の要約がない大量の原文データ、一文の中から不要な情報が除かれた圧縮文を活用して要約モデルの性能向上を行った。具体的には、図8と以下に示すように、主に文法、非冗長性、流暢性、文字数制御の観点で、独自の技術を導入した。その結果、既存技術と比較して高い性能を達成した。

- ①文法誤りのある文を機械的に生成し、文法誤りを低減する強化学習を導入。
- ②冗長な内容が含まれる要約を機械的に生成し、冗長性を低減する対照学習^{*2}を導入。
- ③大量の原文データを用いて流暢性を改善する事前学習を実施。
- ④文字数制御のために圧縮文を用いて学習した文圧縮モデルを導入。

3.2 文法

文法に関しては、特に学習データセットに含まれないような単語の組合せのパターンなどを要約する場合に誤りが起こりやすい。その原因として、文章を生成する深層学習のモデルを学習する際に広く使われている方法である「Teacher forcing」アルゴリズム [1] が挙げられる。深層学習を用いたテキ

原文	要約結果
<p>アメリカカリフォルニア州などで8日午後、大規模な停電が発生した。約300万世帯が影響を受けた他、原子力発電所の原子炉2基が緊急停止している。現地の電力会社によると、停電はサンディエゴなどカリフォルニア州南部とアリゾナ州西部の他、国境を越えたメキシコの一部の地域など広範囲に及び、道路の信号が消えて渋滞が発生するなど約300万世帯が影響を受けたという。この停電により、サンディエゴ近郊のサンオノフレ原発では原子炉2基が緊急停止したが、電力会社は放射能漏れなど、安全に関する問題は起きていないとしている。また、停電の原因について、電力会社はテロではなく変電所の工事に伴うトラブルとの見方を示している。</p>	<p>カリフォルニア州などで8日午後、大規模な停電が発生した。この停電により、サンディエゴ近郊のサンオノフレ原発では原子炉2基が緊急停止したが、電力会社は放射能漏れなど、安全に関する問題は起きていないとしている。</p>

 生成式要約で引用された位置
 単語が新たに生成された位置

図7 要約の引用位置や生成位置の可視化例

^{*2} 対照学習：類似データ間の特徴距離が非類似データ間の特徴距離よりも近くなるように学習することでモデルの高精度化を行う手法。

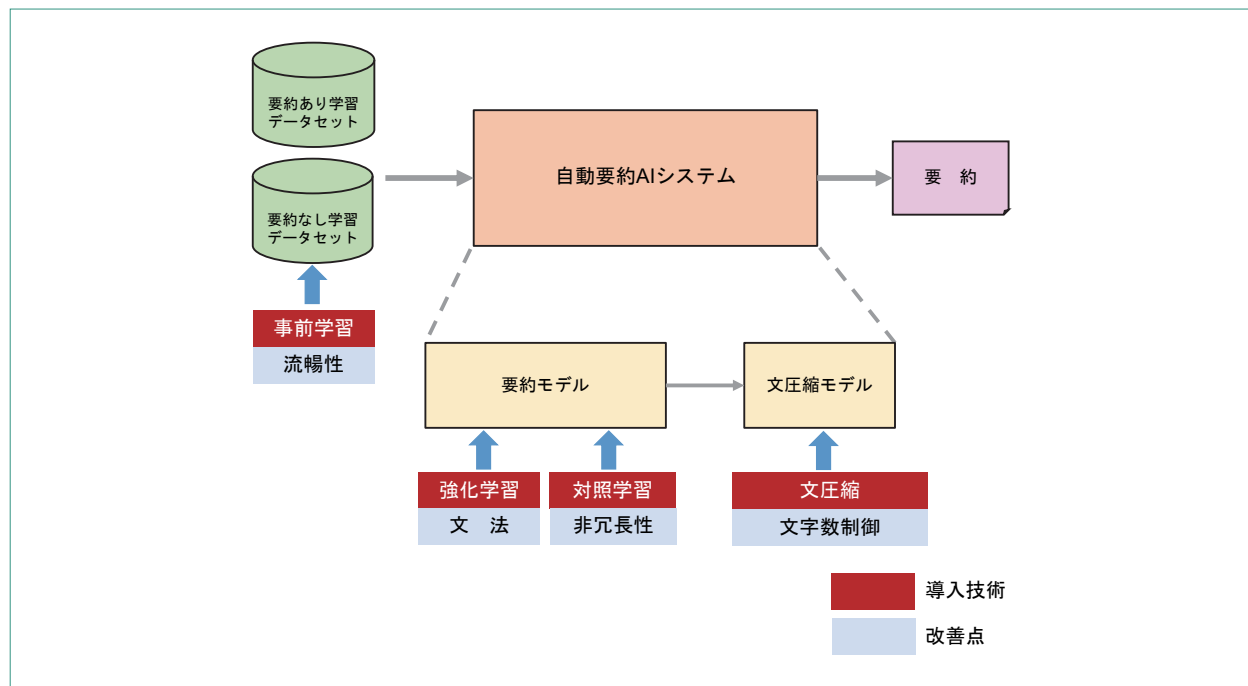


図8 要約システムの改善点

スト生成では、直前に生成した単語を用いて次の単語を生成するが、学習の初期段階では不正解の単語の生成が多いため、Teacher forcingを用いる場合、直前に生成した単語ではなく正解の単語を深層学習のDecoder*3側に入力することで、学習の効率化が行われる。このような学習では、特定の単語に対して次に生成される単語を学習するため、文全体ではなく単語レベルでの局所的な学習が行われる。

課題は、学習時は正解の単語が与えられるにも関わらず、実際に未知の文章を用いてテキスト生成を行う際には、正解の単語ではなく直前の深層学習によって生成された単語を用いてテキストが生成されるため、一度誤りとなる単語が生成されると、その単語の次に生成されるテキストが意味的にも文法的にも破綻しやすいことである。

そこで、ドコモの自動要約AIシステムでは、強化学習の方法を取り入れた。強化学習はエージェントとそれに報酬を返す環境の間のフィードバックを

通して、エージェントを学習させる手法である。一般的にロボット制御などの分野で利用される学習手法であるが、近年自然言語処理の分野でも利用される例がある [2]。本システムでは、強化学習の各要素が下記のように構成される。

- ・ エージェント (Agent)：要約モデル。
- ・ 環境 (Environment)：文を与えたときに文法的に正しいか正しくないかを判断する識別器。
- ・ 状態 (State)：要約結果。
- ・ 動作 (Action)：次の単語の生成。
- ・ 報酬 (Reward)：文法の正しさのスコア（識別器の判断結果）。

要約モデルが要約を生成する行動を通じて、報酬が最も多くなるように、つまり文法の誤りが少なくなるように学習を行うことで、既存手法と比較して文法誤りの少ない要約モデルを構築した。

*3 Decoder：テキストを入力して新たなテキストを出力する多層のニューラルネットワークのうち、テキストを入力に受け取り特徴量化する部分をEncoder。Encoderの出力する特徴を用いて新たなテキストを生成する部分をDecoderと呼ぶ。

3.3 非冗長性

本稿では、要約内容の中で、同一文内や複数の文間で意味的に同一の内容が繰り返し言及されることを「冗長」と表現する。深層学習を用いたテキスト生成モデルでは、冗長な内容を含む要約の生成が課題として指摘されている [3]。

そこで、ドコモの自動要約AIシステムでは、対照学習を導入することにより本課題の改善を行った。対照学習では、Anchor、正例、負例の3つを用意し、学習時にAnchorと正例の特徴間の距離が、Anchorと負例の特徴間の距離よりも近接するように学習することでモデルの精度を向上する手法である。ドコモの自動要約AIシステムでは、Anchorは要約モデルによって生成された要約、正例は人手で作成した正しい要約であり、負例は正例に対して機械的に同一の単語、フレーズ、文を繰り返した、冗長性の観点で誤りのある要約である。このように冗長性の観点で誤りのある負例を用いた学習を行うことで、既存手法と比較して冗長な文を生成しにくい要約モデルを構築した。

3.4 流暢性

本稿では、単語やフレーズ間の組合せが正しいことを「流暢」と表現する。実際のシステム利用時に、学習データセットの中で見られない単語の組合せを含む原文が入力される場合がある。このとき既存手法を用いると、文法的に正しくても使い方の誤った単語の組合せで要約が出力されることがある。

このような課題を解決するために、ドコモの自動要約AIシステムでは、大量の文章を用いて単語の組合せに関する事前学習を行い、要約モデルの流暢性の向上を図った。

3.5 文字数制御

文字数制御で最も一般的な手法は、要約を生成する過程でBeam Search^{*4}を行う際に、ルールベース

でより適切な長さとなる出力結果を選択する手法である。このようなルールベースの手法の場合、適切な長さとなる出力結果を選択するために、文法的な正しさや主題の抽出度合いについて考慮されない場合があり、要約結果がユーザの指定した長さに近い一方で、文法的に誤りのある要約や主題から逸脱した要約が出力されやすくなる [4]。

そこでドコモの文字数制御機能は、要約モデル自体に出力する要約の長さの情報を特徴量^{*5}として入力し、要約モデルが要約を出力する際に文法、主題、長さなどを同時に考慮するように学習させることで、要約モデルの最適化を行った。また要約モデルの後処理として、文字数を削減するための文圧縮モデルを導入することにより、要約モデルより出力された文がユーザの設定する文字数を上回る場合は、文を適切な長さに圧縮し要約を生成できるように改善した。

4. 生成式要約の性能評価

4.1 データセット

日本テレビ放送網株式会社が提供する約18万件のニュース記事とそれらを人手で要約したデータセットを利用し、本システムの生成式要約の学習と評価を行った。

4.2 ROUGEによる評価

正解要約 (Reference) に対する要約結果の網羅性を評価するRecallについて各ROUGE (Recall-Oriented Understudy for Gisting Evaluation) の計算方法を図9に示す。ROUGEは、正解とされるテキストとモデルが作成したテキストの類似度を比較する要約の指標の中で最も広く利用されるものである [5]。ROUGE-1、ROUGE-2は、それぞれテキスト間のunigram^{*6}とbigram^{*7}の重なり度合いを表し、ROUGE-Lは、一致する最長の単語の長さを用いてテキスト間の重なり度合いを測定する。ROUGE-1、

*4 Beam Search：本稿では、ニューラルネットワークが出力する単語の候補をスコアに基づいて複数選定し、いくつかの要約結果の候補を得ることを指す。

*5 特徴量：データから抽出される、そのデータを特徴付ける量 (数値) のこと。

*6 unigram： n 単語連続して続く文字列を n -gram と呼び、 n が1の場合の1単語だけの文字列を指す。

*7 bigram：2単語連続して続く文字列をbigramと呼ぶ。

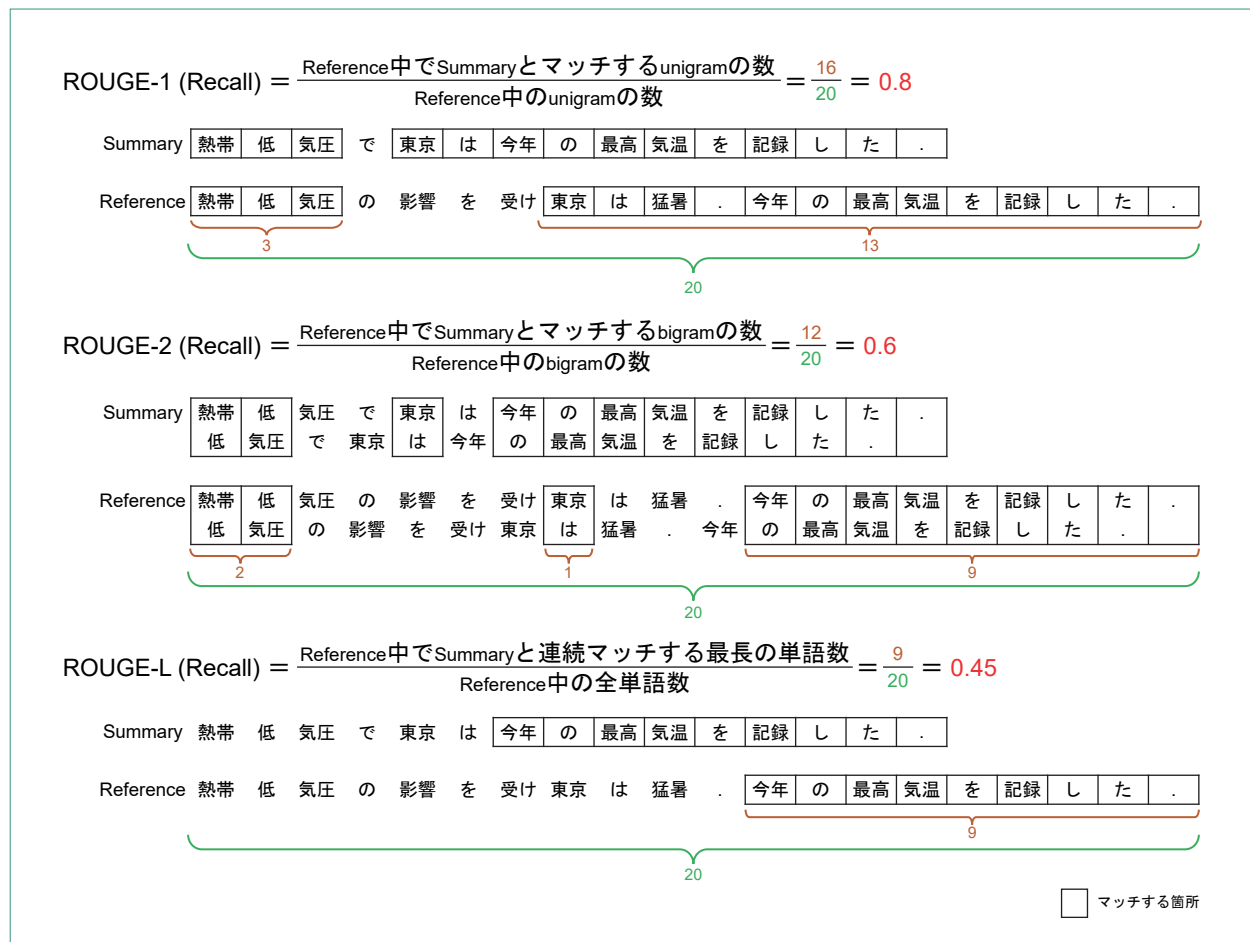


図9 各ROUGEの計算方法

ROUGE-2, ROUGE-Lすべてにおいて、値が大きいほどテキスト間の重なりが大きく、テキスト生成モデルの性能が高いことを示す。

表1は、日本テレビ放送網株式会社が提供するデータセットのうち3,000件を用いて、本システムの評価を行った結果である。比較対象として原文の先頭3つの文を要約結果とするLead-3、また、TextRank [6], SumBasic [7], LSA (Latent Semantic Analysis) [8], Submodular [9], PGN (Pointer Generator Network) [3] を利用した。TextRank, SumBasic, LSA, Submodularについては抽出式要約であり、抽出する文の中で重要度の高い3つの文

を要約結果とした。PGNについては生成式要約であり、Beam Search結果の中で最も正解の要約の文字数に近い要約を、最終的な要約結果として評価した。これらの既存手法と比較して、ドコモの自動要約AIシステムはROUGEの数値が大きく、人手で作成した正解に、より近い要約を生成することが示された。

4.3 人手による評価

前述のROUGEの評価による欠点として、文法の誤りや意味的な冗長性などを評価できない点が挙げられる。そこで、日本語を母国語とする者が以下の

4つの観点で評価を行った。その際スコアを、4を最高点とする4段階評価とした。

- ①文法：要約に文法的な誤りが少ないこと。
- ②主題性：生成された要約が原文の主要な内容をカバーしていること。
- ③非冗長性：要約の中に意味的に同一の単語、フレーズ、文の繰り返しが無いこと。
- ④流暢性：生成された要約が単語間、文間で流暢であること。

表2は、評価データセットのうち100件を用いてドコモの自動要約AIシステムとPGNの出力結果を人手で評価した評価値の平均値である。また表3に示すように、指定した要約の文字数に対して生成された要約の文字数の割合を算出し、4段階のスコアに割り当てた。この要約文字数のスコアを評価データごとに算出し、平均値を表2の「要約文字数」列に記載した。このような人手による評価についても既存手法のPGNと比較して、ドコモの自動要約AI

システムは各評価指標の数値が大きく、より品質の高い要約が生成されることが示された。なお、自動要約の処理時間は、抽出式要約が約1秒、生成式要約が10秒程度のため、数分から数十分かかる手動の要約作業と比較して速度が大幅に向上する。

5. あとがき

本稿では、ドコモの自動要約AIシステムの各機能や技術、性能について解説した。ドコモは、ユーザがヒントを指定するヒント機能、タイトルを使用するタイトル機能、要約の文字数や文の数を制御する文字数制御機能、原文上の引用位置や新たに生成された文字の位置を可視化する位置特定機能を開発し、ユーザの意向に沿った要約を出力しやすいシステムを実現した。また、文法、非冗長性、流暢性、要約文字数の制約の課題を解決するために、強化学習、対照学習、事前学習などを導入し、性能向上を行った。性能評価結果が示すように、ドコモの自動

表1 ROUGEによる評価結果

		ROUGE-1	ROUGE-2	ROUGE-L
抽出式要約	Lead-3	74.46	63.89	72.48
	TextRank	64.06	50.07	60.16
	SumBasic	64.49	49.18	58.38
	LSA	62.28	46.48	56.85
	Submodular	55.41	36.91	47.41
生成式要約	PGN	79.25	70.36	77.45
	本システム	84.49	76.47	81.80

表2 人手による評価結果

	文法	主題性	非冗長性	流暢性	要約文字数
PGN	2.82	2.40	3.89	3.06	3.25
本システム	3.85	3.53	3.92	3.84	3.89

表3 要約の文字数に関するスコア

要約の文字数範囲	スコア
$0.7 \times L \leq S \leq 1.0 \times L$	4
$0.6 \times L \leq S < 0.7 \times L$ or $1.0 \times L < S \leq 1.1 \times L$	3
$0.5 \times L \leq S < 0.6 \times L$ or $1.1 \times L < S \leq 1.2 \times L$	2
$S < 0.5 \times L$ or $1.2 \times L < S$	1

L : 指定した要約の文字数

S : 要約モデルが作成した要約の文字数

要約AIシステムは既存の技術と比較し、ROUGEによる評価と人手による評価の両方の指標の値を大幅に改善した。この自動要約AIシステムにより、要約作業時間を短縮し人手不足を解消することが可能となる。今後は、実サービスを通して得られた課題を基に既存機能の性能向上や新機能の開発を行い、さらに高性能な自動要約AIシステムを実現していく。

文献

- [1] R. J. Williams and D. Zipser : "A Learning Algorithm for Continually Running Fully Recurrent Neural Networks," Neural Comput., Vol.1, No.2, pp.270-280, 1989 (doi : 10.1162/neco.1989.1.2.270).
- [2] L. Yu, W. Zhang, J. Wang and Y. Yu : "SeqGAN : Sequence generative adversarial nets with policy gradient," 31st AAAI Conf. Artif. on Intell., pp.2852-2858, 2017.
- [3] A. See, P. J. Liu, and C. D. Manning : "Get To The Point : Summarization with Pointer-Generator Networks," ACL 2017 - 55th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.), Vol.1, pp.1073-1083, 2017(doi : 10.18653/v1/P17-1099).
- [4] B. Eikema and W. Aziz : "Is MAP Decoding All You Need? The Inadequacy of the Mode in Neural Machine Translation," Proc. of the 28th International Conference on Computational Linguistics, pp.4506-4520, 2021 (doi : 10.18653/v1/2020.coling-main.398).
- [5] C.-Y. Lin : "Looking for a Few Good Metrics : ROUGE and its Evaluation," NTCIR Work., pp.1-8, Jun. 2004.
- [6] R. Mihalcea and P. Tarau : "TextRank : Bringing Order into Text," Proc. of 2004 Conf. Empir. methods Nat. Lang. Process., pp. 404-411, 2004.
- [7] L. Vanderwende, H. Suzuki, C. Brockett and A. Nenkova : "Beyond SumBasic : Task-focused summarization with sentence simplification and lexical expansion," Inf. Process. Manag., Vol.43, No.6, pp.1606-1618, NoV. 2007 (doi : 10.1016/j.ipm.2007.01.023).
- [8] M. G. Ozsoy, F. N. Alpaslan and I. Cicekli : "Text summarization using latent semantic analysis," J. Inf. Sci., Vol.37, No.4, pp.405-417, Aug. 2011 (doi : 10.1177/0165551511408848).
- [9] H. Lin and J. Bilmes : "A class of submodular functions for document summarization," Proc. of 49th Annu. Meet. Assoc. Comput. Linguist. Hum. Lang. Technol., Vol.1, pp.510-520, Jun. 2011.