

生成型自動要約の信頼性向上を目的とした 数値情報の誤り検出と修正手法

松井我颯¹ 石川晴基² 中島陽子³ 本間宏利³ 秋葉友良²

¹ 釧路工業高等専門学校 電子情報システム工学専攻2年 ² 豊橋技術科学大学

³ 釧路工業高等専門学校 創造工学科

s230710@kushiro.kosen-ac.jp, ishikawa.haruki.yu@tut.jp,

{yoko, honma}@kushiro-ct.ac.jp, akiba@cs.tut.ac.jp

概要

自動要約技術は地方議会の議事録やネットニュースなどの情報の処理に有用だが、生成型要約において数値情報の誤りが問題となっている。本研究は、自動要約における数値情報の正確性を向上させるため、松井らが提案した二つの手法であるセグメントの特定による数値探索手法 (SSR: SegmentSummaryRevision) と係り受け解析に基づく数値探索手法 (DSR: Dependency-basedSummaryRevision) を組み合わせた手法を提案する。提案手法では、まず原文から要約に用いられている数値が含まれる文を特定し、次に用いて特定した文と要約文の数値を文脈に基づき比較する。この一連の処理により、要約文中の数値を識別し、修正する。

1 はじめに

近年、政治関連のフェイクニュースが世界的な社会問題になっている。誤情報の拡散は公共の認識を歪め、混乱を招く恐れがある。この問題に対処するために、多くの研究者が自然言語処理技術を用いた自動要約の応用に注目している。現在主流の生成型自動要約技術は、大量のテキストデータから情報を学習し、簡潔に表現することで、情報の迅速な理解と判断を支援することができる [1]。しかし、小川ら [2] も指摘しているように、生成型自動要約技術には正確な数値情報の取得に失敗し、不正確な数値で要約を行ってしまうケースが見られる。数値データは重要な情報であるため、このような誤りは、政治関連のニュースにおいて特に重要な問題となり得る。この問題に対して松井らは自動要約に含まれる数値情報の正確性を原文との比較に基づいて評価する二手法を提案している [3] [4]。これらの手法では、文章の構造と文脈という異なる観点か

ら数値の正誤判定を行う。本研究では、それらの手法を組み合わせたとともに、誤りと判定した要約の数値を修正をする機構を構築し、評価を行う。対象とする要約データは、都議会議事録における答弁とその自動要約文である。

2 テストデータ

本研究では NTCIR-17QALab-PoliInfo-4 の AnswerVerification タスクで配布されたトレーニングデータの一部である都議会議事録の答弁の全文 (AnswerOriginal), 答弁の要約 (AnswerSummary), 会議が行われた日付 (Date), 要約の正誤 (PredictedClass) を用いる [5]。表 1 に、答弁の全文と答弁の要約の例を示す。

2.1 前処理

AnswerOriginal と AnswerSummary のテキストについて、自動要約時に一部の表現が変更されることがあるため、これらの表現を統一する目的で以下の項目 a, b, c に示す 3 つの前処理を行う。前処理を施した AnswerOriginal と AnswerSummary をそれぞれ原文と要約文とする。

a. 和暦を西暦に変換

原文では、年についての言及時に「平成 29 年」「令和 2 年」といった和暦が用いられているが、これにより年号を考慮する必要が生じ、数値の正誤判定が困難になる。そこで、判定を容易にするために、すべて西暦表記に統一する。また、原文で「今年」などと記載されている場合、要約では「令和 2 年」などの数値で表記されている場合がある。このため、Date を参照し、「今年」「元年度」などの表現も西暦に変換する。

表 1: 答弁と要約文の例

答弁の全文	<p>まず、インクルーシブな公園整備についてのお尋ねでございます。誰もが自分らしく輝くことのできるダイバーシティの実現に向けて、都立公園において、障害の有無にかかわらず全ての子供たちが安全に楽しむことのできる遊び場、これを整備することは重要でございます。都といたしまして、今年度、障害児の保護者、そして障害者団体、障害児保育の現場、ユニバーサルデザインの有識者など、さまざまな方々にヒアリングを行ってまいりました。その中で、体を支える力が弱い子供さんたちが揺れる感覚を楽しめるそんな遊具や直射日光を避けることのできる休憩場所の設置など、さまざまなご意見をいただいたところでございます。こうしたご意見を踏まえまして、現在、砧公園と府中の森公園を対象に、具体的な設計を行っておりまして、平成三十一年度末の完成を目指し整備を進めてまいります。今後とも、都立公園でこうした取り組みを進めていくことで、障害の有無にかかわらず、全ての子供たちがともに遊び、また、学ぶ機会を積極的に提供してまいります。</p>
答弁の要約	<p>障害者団体等にヒアリング。砧公園と府中の森公園を対象に 31 年度末完成を目指す。</p>

b. 漢数字をアラビア数字に変換

数値による判定を容易にするために、漢数字をアラビア数字に変換する。この変換では、まず GiNZA¹を用いて分かち書きを行い、数値を抽出する [6]。モデルは ja_ginza_electra²を用いる。その後、抽出された数値を kanjize³を用いて変換する。また、今回採用しているモデルでは「一回」や「一人」といった漢数字の「一」が含まれる単語は合成語として一つの単語として扱われるため、個別でアラビア数字に変換するよう対応する。

c. 数値を使用した要約の抽出

本実験は数値の誤りに着目するため、正規表現を用いて数値が含まれている要約文をもつデータの組のみ抽出してテストデータとする。

3 提案手法

松井らが提案した手法のうち、第一の手法、セグメントの特定による数値探索手法 (SSR:SegmentSummaryRevision) は、原文から要約に用いられている数値が含まれる文 (キーセグメント) を特定する。これにより、文章内に複数の数値が含まれている場合でも、要約文の正誤判定

¹<https://megagonlabs.github.io/ginza/>

²<https://huggingface.co/megagonlabs/transformers-ud-japanese-electra-base-discriminator/>

³<https://pypi.org/project/kanjize/>

に必要な数値のみ取得することが可能である。二つ目の手法、係り受け解析に基づく数値探索手法 (DSR:Dependency-basedSummaryRevision) は、各数値の係り受け元を特定し、その情報を基に数値を比較する。この比較により、文脈上の誤りや、数値の不適切な利用を識別し、修正することが可能である。

本研究では SSR と DSR を組み合わせる。始めにキーセグメントを特定し、その後要約文とキーセグメントに含まれる各数値の係り受け元を特定し、その情報を基に数値を比較する。この比較により、数値の不適切な利用を識別し、修正することが可能である。

3.1 キーセグメントの特定

原文から、要約に使用されている数値と単位が含まれる文を特定する。要約内容が含まれる箇所を特定するため、まず要約元の文章を句点 (。) を基準に分割し、それによって得られる各文を個別のセグメントとして扱う (表 2)。要約に用いられた特定のセグメントを識別するために、TermFrequency-InverseDocumentFrequency (Tf-Idf) を採用する。Tf-Idf は、文書内の各単語の重要度を測定する手法であり、文書全体の中で特定の単語がどの程度重要であるかを定量的に評価することが可能である。

算出した Tf-Idf 値に基づいて、要約文と各セグメント間の類似度の評価を行う。この評価は、各セグメントにおける単語の Tf-Idf 値から構成されるベクトルを基に、コサイン類似度を計算することで行う。コサイン類似度は、二つのベクトル間の角度のコサインを計算することで、これらのベクトルがどの程度似ているかを測定する手法である。具体的には、各セグメントにおける単語の Tf-Idf 値をベクトルとして表現し、要約文と各セグメントのベクトル間でコサイン類似度を計算する。要約文と各セグメント間のコサイン類似度を比較し、最も高い値を持つセグメントをキーセグメントとして抽出する。これにより、要約文に最も類似するセグメントを取得し、正誤判定に必要な数値を効率的に探索可能となる。

3.2 係り受け解析に基づく数値探索

キーセグメントと要約文に対して係り受け解析を用い、数値がどの事柄に関するのかを特定し、数値の比較を行う。係り受け解析器には GiNZA を用いる。キーセグメントと要約文内のトークン (単語や記号) に対して係り受け解析を行うことにより、各トークン間に付与される依存関係ラベルを用いて、主語や述語などの文法的な関係を把握することが可能となる。また、

表 2: セグメント分割の例

セグメント 1	都立学校における普通教室の無線 LAN 整備率は、2019 年 3 月時点で約 16%でございまして、BYOD 研究校の成果を全都立学校に広げていくためには、通信環境整備の加速が必要でございます。
セグメント 2	そのため、2020 年度に都教育委員会は、都立学校 80 校の普通教室、特別教室等に無線 LAN 環境を整備いたします。
セグメント 3	こうした取り組みによりまして、スマート東京実施戦略が掲げる全ての児童生徒がインターネットにつながる環境の実現を目指してまいります。

各トークンに付与される品詞ラベルを用いて単語の品詞を把握することが可能となる。

要約文から数値に関するトークン(キートークン)を取得するために、係り受け解析結果を用いて数詞の位置を特定する。文頭から、数詞を意味する品詞ラベルである NUM ラベルが付与されているトークン(NUM トークン)を探索し、これを開始位置とする。次に、NUM トークンに数詞と名詞の関係を表す依存関係ラベルである nummod ラベルが付与されている場合、係元をその数値の単位として取得する。要約文から単位の係元にあたるトークンの原形を取得し、キートークンとする(図 1)。また、図 2 のように nummod ラベルではなく、複合名詞を意味する compound ラベルが付与されている場合がある。この場合は、NUM トークンの次に格納されているトークンと、その係元のトークンを接続し、単位として取得する。図 2 では、これが「年」と「末」を結合した「年末」にあたる。また、修正時に必要なため、数詞の位置を同時に取得する。この位置情報は、文中において NUM トークンが何番目に出現したかを示すものである。

その後、キーセグメントからキートークンを探索し、その係先にあたる数値と単位を取得する。これらは、数値、単位、キートークンを三つ組みと(e.g.[2018, 年, 発送])して格納したリストを生成する。キーセグメントと要約文に対して適用した例を表 3 に示す。

表 3: 取得した数値とキートークンのリストの例

	テキストデータ	リスト
原文	そのため、 2020 年度 に都教育委員会は、都立学校 80 校 の普通教室、特別教室等に無線 LAN 環境を 整備 いたします。	[[2020, 年度, 整備], [80, 校, 教室]]
要約文	2020 年度 に都立学校 810 校 の普通教室、特別教室等に 整備 。	[[2020, 年度, 整備, 0], [810, 校, 教室, 5]]

3.3 要約の正誤判定

キーセグメントと要約文のリストを比較し、同じキートークンの数値が一致しなければ、要約文の数値を誤りと判定する。例えば、表 3 のようにキートークン「教室」の係り先である「80 校」と「810 校」の数値が異なる場合、要約文中の「810 校」を誤りと判定する。

3.4 要約の修正機構

数値の修正は、要約文のリストに含まれている数値の位置情報を参照し、誤っている数値を原文の正しい数値に置き換えて行う。正しい数値はキーセグメントから生成された三つ組みのリストに含まれる数値を用いる。修正例として、表 3 の誤要約を修正した要約文を表 4 に示す。

表 4: 表 3 の要約文修正の例

修正前	2020 年度に都立学校 810 校の普通教室、特別教室等に整備。
修正後	2020 年度に都立学校 80 校の普通教室、特別教室等に整備。

4 評価実験

提案手法の有効性を検証するため、評価実験を行う。実験にはテストデータから、数値が含まれていて正しい要約 75 件、数値が誤っている要約 35 件、計 110 件用いる。実験は要約の正誤判定を行う機構と、セクション 3.4 で説明した要約修正機構に分けて行う。

4.1 正誤判定機構の実験設定

原文と要約文に対して本手法を適用する。評価指標は、テストデータに付与されている自動要約の正誤(PredictedClass) タグを基準に、要約の正誤を正確に判定した割合を示す正解率、正しい要約に対する正答率を示す再現率、誤要約を誤りと正しく判定した割合を示す特異度を算出し、提案手法の有用性を検証する。特異度(S)は、真陰性を陰性の総数で除した値として式(1)で表される。各評価指標は 0.0 から 1.0 までの値を取る。

$$S = \frac{\text{真陰性}}{\text{真陰性} + \text{偽陽性}} \quad (1)$$

本実験では、全体の予測がどの程度正しいか判断するため正解率を評価指標として用いる。また、正しい要約文を正しい、誤っている要約文を誤りと判定する

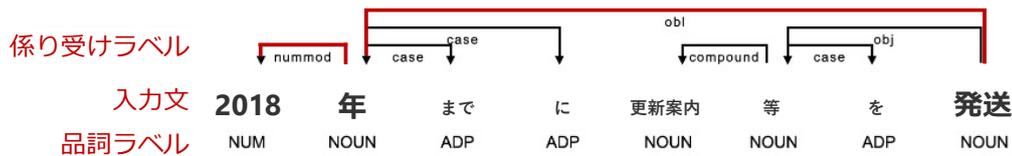


図 1: 係り受けラベルと品詞ラベルの例 1

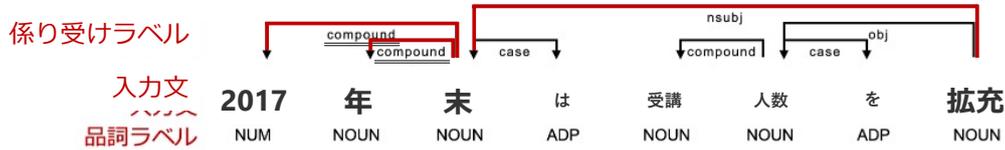


図 2: 係り受けラベルと品詞ラベルの例 2

ことが重要であるため、再現率と特異度も指標として用いる。

4.2 正誤判定機構の実験結果

実験結果を表 5 に示す。本手法の正誤判定の正解率は 0.727 であった。しかし、再現率は 0.600 であり、正しい要約を誤りであると判定する場合がある。特異度は 1.000 であるが、キーセグメントにキートークンが含まれておらず、誤りと判定しているが正しい数値の比較が行えていない場合がある。松井らの DSR と比較すると、正解率が 0.027 低下している。

正誤判定に失敗した例として、複数の数値に同じキートークンが関連している場合、原文から正しく数値を取得できず、判定に失敗する場合がある。表 6 に示す例では、原文から「4」と「20」という数値が取得されるべきである。しかし、要約文から取得した数値に対応するキートークンがいずれも「拡充」となっていたため、「20」ではなく誤って「4」を取得している。このような場合では、数値情報が正しい要約を誤りであると判定される。

表 5: 評価実験の結果

手法	正解率	再現率	特異度
SSR	0.573	0.387	0.971
DSR	0.754	0.667	0.914
本手法	0.727	0.600	1.000

4.3 要約修正機構の評価実験

要約修正機構についての評価実験を行う。正誤判定機構で数値が正しく取得で来ている誤要約 20 文に対

表 6: 正誤判定の失敗例

	テキストデータ	三つ組みリスト
要約文	職員宿舍借り上げへの補助上限戸数を 4 戸から利用定員に応じて最大 20 戸に拡充。新規受け付け期間も延長し事業者の取組を支援。	[[4, 戸, 拡充, 9], [20, 戸, 拡充, 18]]
原文	(前略) 災害時における福祉避難所の運営体制の強化とを目的とした、職員宿舍借り上げへの補助の上限戸数を 4 戸から利用定員に応じて最大 20 戸に拡充するほか、新規受け付け期間も令和 5 年度まで 3 年間延長しておりまして、今後、介護職員の育成、定着に向けた事業者の取り組みを支援してまいります。	[[4, 戸, 拡充], [4, 戸, 拡充]]

して修正手法を適用し、正しい数値に修正されているか手動で確認を行う。実験の結果、数値は 100% の精度で修正された。これにより、提案する要約修正機構が、数値が間違っていると判定された要約文について、数値の取得が正しく行われている場合には適切に修正可能であることが示された。

5 おわりに

本研究では、都議会議事録の答弁とその自動要約文を対象として、原文を参照して要約の数値情報の正誤を判定し、誤っていた場合に数値を修正する手法を提案した。複数の数値が同一のキートークンを係元としている場合やキーセグメントにキートークンが含まれていない場合に対応するため、今後は、係り受け解析木のマッチングにより文全体の文脈を考慮することで改善を目指す。

参考文献

- [1] 石垣達也, 高村大也, and 奥村学. ”複数文質問を対象とした抽出型および生成型要約.” 自然言語処理 26.1 (2019): 37-58.
- [2] Ogawa, Yasuhiro, Yugo Kato, and Katsuhiko Toyama. ”nukl ’ s QA System at the NTCIR-16 QA Lab-PoliInfo-3.” Proceedings of The 16th NTCIR Conference. 2022.
- [3] 松井我颯, 石川晴基, 中島陽子, 本間宏利, 秋葉友良, “都議会議事録における自動要約のための数値情報自動修正手法の提案,” in Proc. 30th Annu. Meeting Assoc. Lang. Process., 2024.
- [4] Gafu Matsui, Haruki Ishikawa, Yoko Nakajima, Hiroto Honma, Tomoyosi Akiba, ”Proposal for Automatic Numerical Correction in Meeting Record Summarization,” The 9th International Conference on Science of Technology Innovation (STI-Gigaku2024), STI-9-61, Nagaoka, Nov., 2024.
- [5] 浅原正幸, 金山博, 宮尾祐介, 田中貴秋, 大村舞, 村脇有吾, 松本裕治. ”Universal Dependencies 日本語コーパス.” 自然言語処理 26.1 (2019): 3-36.
- [6] Yasuhiro Ogawa, Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ototake, Yuzu Uchida, Keiichi Takamaru, Kazuma Kadowaki, Tomoyoshi Akiba, Minoru Sasaki, and Akio Kobayashi. Overview of the NTCIR-17 QA Lab-PoliInfo-4 Task. In Proceedings of the 17th NTCIR Conference, 2023.