

答弁の種類に着目した抽象型要約に基づく議会会議録質問応答

河合輝也¹ 秋葉友良¹ 増山繁²

¹ 豊橋技術科学大学 ² 東京理科大学

kawai.teruya.ic@tut.jp akiba@cs.tut.ac.jp masuyama@rs.tut.ac.jp

概要

地方議会がウェブで公開している議会会議録は、長く複雑な構造をしているため読みづらく十分に活用されているとは言い難い。本研究では、住民が議事録を読まずに行政の疑問を解決できる質問応答手法を提案する。特に、対話、ニュース記事を複数の側面から要約する研究を、本タスクに適用する。まず、答弁要約のクラスを4種類定義し、サンプルから分布を調べた。その結果、ある議案に関して既に行った対策・措置である「進捗」、これから行う対策・措置である「計画」が全体の92%を占めていることを確認した。質問に対する回答を「進捗」と「計画」に着目して要約するモデルを実装し、ベースラインを大幅に上回る性能を示した。

1 はじめに

近年、地方議会が情報を一般公開し行政の透明化を進めている。すべての都道府県議会が独自に情報公開の条例を新たに制定・改定し、現在は本会議の会議録がインターネットで閲覧できる。本会議では委員会の審議を通過した議案について質疑討論が行われる。住民は会議の概要を「～だより」のような刊行物や各種報道を通じて知る。これらのメディアを通じて得られる情報は、多くの住民に影響があるものや話題性が高いものが中心で議論すべてを網羅していない。ごく少数の住民に関係のある細かい内容は報道されない場合も多く、詳細を知りたい住民は最終的に公開されている会議録を読む必要がある。

議会会議録は複雑な構造で長く読みづらいという問題がある。例えば、東京都議会の会議録は一括質問一括答弁形式になっている。ある場合には議員が約2万2千文字で7つ質問した後に、知事が約2万3千文字で7つ答弁している。このように、挨拶や議論とは直接関係ない話を含み長くなることと、一括質問・一括答弁で質問と答弁の対応関係が複雑な

ため、読みづらい。

本研究では、議会会議録を基に行政に関する質問の回答を出力する質問応答を提案する。提案手法では質問に関連する答弁を抽出・要約する際、特に答弁の種類に着目して設計した。東京都の職員が作成した会議録の人手要約を分析し、4種類にクラス定義、分類した。4クラス中の「進捗」と「計画」が要約生成において重要な要素であるという仮説を立て、モデルを実装した。

2 関連研究

質問応答(QA)システムは、一般的に質問分析、情報検索、回答抽出、回答選択の各要素から構成される。初期の研究では、ルールベースや統計的手法が用いられていたが、現在ではほとんどの研究がニューラルネットワークを使用している。[1, 2, 3] QAシステムが扱う質問は、ファクトイド質問とノンファクトイド質問に分けることができる。ファクトイド質問とは、名前、番号、場所など、事実に基づいた答えを必要とする質問である。非ファクトイド質問とは、理由や事象の説明に基づく回答を求める質問である。QAシステムの研究では、ファクトイド型質問が主流だが、この研究ではより難易度の高いノンファクトイド型に取り組む。

テキストの要約は、ニュース記事、論文、書籍など、様々な分野で研究されている。要約には、抽出型と抽象型の2種類がある。前者はテキストから単語を選択して文章を生成するものであり、後者は単語の意味情報を捉えて文章を生成するものである。抽出的要約の研究として、Kevin Knight と Daniel Marcu [4] は、文章を構造的に分析し、木構造にして重要な情報を圧縮する方法を提案した。TextRank[5] は文章をグラフ構造で表し、各ノード間の関係から要約を作成する。Latent Semantic Analysis[6] を用いたテキスト要約は、文の行列の特異値分解によりトピックを計算し、文を選択するトピックベースの手法である。

最近の研究で注目されているのが抽象的要約である。Encoder-Decoder モデルを用いて、入力に存在しない単語を出力することが可能になった [7]。これは、Encoder が文を潜在的な表現に圧縮し、Decoder が圧縮された表現から文を生成するモデルである。抽象型要約は文の潜在的な意味を捉えた表現を可能にするが、しばしば冗長な情報や矛盾した情報を生成してしまうという問題がある。Pointer-Generator Networks は、原文から単語をコピーできるアーキテクチャを提案した [8]。また、Transformer は、RNN や CNN を用いず、Attention のみを用いた高速かつ高精度な Encoder-Decoder モデルである [9]。

対話要約タスクを Chen らはマルチビュー Sequence-to-Sequence モデルで取り組んだ [10]。分類モデルで会話構造を明確にし、対話要約の複数のビューで生成した。ニュース記事のテキストをドメインに合わせた複数の側面 (アスペクト) から要約する研究がある [11]。クエリベースの要約は文書固有の質問に答えるが、実際のユーザーのニーズは多くの場合、抽象的な疑問の解決であると著者らは考えている。ニュース記事の属するドメインにあらかじめ定義された複数の側面から要約を生成する。例えば、地震ドメインのニュースは震源地、震度などの地理的側面と被災者の状況、政府の対応などの被害・復興的側面から要約するように多様なニーズに応える要約生成を著者らは提案した。

3 提案手法

提案手法を図 1 に示す。提案手法は抽出・要約フローをコンテンツ選択器と要約モデルで実装している。コンテンツ選択で答弁の「進捗」と「計画」を抽出する。それぞれを要約モデルに入力し、進捗要約と計画要約を生成する。

東京都議会は会議録を職員が要約し「都議会だより」でウェブで公開している。人が要約を作成する際どのような側面に着目するか調べるため、都議会だよりの答弁要約を分類した。人手答弁要約は基本的に 1~2 文で構成されている。119 件の答弁要約を文集合に分割し、答弁要約を分類した。答弁要約のクラス定義とクラス分類結果を表 1, 2 に示す。進捗・計画と分類されたものが全体の約 92% を占める。よって、その 2 つの側面から要約を生成することを今回の方針とした。

表 1 答弁要約のクラス定義

進捗	プロジェクトの進み具合。 すでに行った対策・措置。
計画	プロジェクトの今後の方針。 これから行う対策・措置。
見解	物事の主観的評価・意見。
事実	現実起きた事象。 定量的データ。

表 2 答弁要約のクラス分類結果

進捗	計画	見解	事実
22	87	7	3

3.1 コンテンツ選択器

都議会会議録は一括質問一括答弁形式で記述されているため、1 発言が複数のトピック (話題) で構成されている。質問に関する答弁部分を抽出するため、答弁をトピック毎のセグメントに分割する。Kanasaki ら [12] の研究を参考に、接続詞、開始表現、終了表現の辞書を作成、マッチしたテキストの前後で分割する。答弁をセグメント集合 $S = \{s_1, s_2, \dots, s_n\}$ (n : トピック数) と表す。

次に、クエリ (質問とトピック) に基づいてセグメント集合から 1 つのセグメントを選択する。セグメントの検索には統計的手法である Okapi BM25 [13] を用いる。提案手法におけるクエリは、質問とそれに関するトピックである。セグメントを検索する際に、質問とトピックは重要性が異なると考えた。質問 q とトピック t のスコアをそれぞれ BM25 で計算し、重み α でバランスをとる、式 1 のように設計した。

$$\bar{s} = \arg \max_{s_i \in S} (\alpha \text{score}(s_i, q) + (1 - \alpha) \text{score}(s_i, t)) \quad (1)$$

予備実験により、 $\alpha = 0.4$ のときセグメント選択のパフォーマンスが最大になることが分かった。また、セグメント選択手法は、Okapi BM25 と Sentence BERT [14] を比較した結果、今回のタスクでは前者の手法が優れていたためこちらを採用した。

3.2 要約モデル

選択したセグメントからニューラルネット要約モデルで要約を生成する。進捗・計画要約を 2 モデルで行う「進捗/計画要約 2 モデル」と 1 モデルで行う「進捗/計画要約 1 モデル」の 2 つの手法を試した。図 2 に 2 つの要約手法を示す。2 モデルは進捗要約モデルと計画要約モデルが独立して存在する。それぞれにセグメントを入力すると進捗要約、計画要約

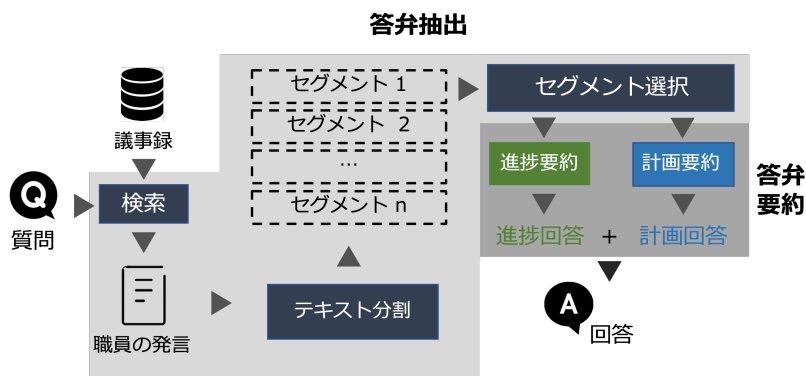


図1 提案手法

●進捗/計画要約2モデル



●進捗/計画要約1モデル



図2 2つの要約手法

が出力される。1モデルは共通の進捗/計画要約モデルで進捗要約、計画要約を生成する。入力セグメントにタグ付けすることで両者を区別し、進捗タグには進捗要約、計画タグには計画要約が出力される。1モデルは共通のモデルによってデータセットが互いに共有され、要約の性能向上に寄与することを期待した。進捗/計画要約2モデル、進捗/計画要約1モデルのいずれもセグメントを進捗・計画の側面から要約する進捗要約、計画要約を生成する。最終的に進捗要約と計画要約を結合したテキストを質問に関する回答(要約)とする。

どちらの手法も、モデル学習のために答弁セグメント-進捗要約、答弁セグメント-計画要約のソース・ターゲットペアが必要である。都議会だよりの人手要約には進捗、要約の区別がない。ターゲット側のデータ構築のため、人手要約を進捗要約・計画要約に分けるニューラルネット分類モデルを使用した。人手要約は文単位に分割し、学習済みクラス分類モデルによって進捗、計画、その他に分けられる。進捗と計画に分類された文をそれぞれ進捗要約モデル、計画要約モデルのターゲットとして学習に利用する。分類モデルは、表1のクラス定義に従って478件の人出要約のラベル付けデータを作成し、学習させた。

4 実験

表3 会議録のデータ構造

Volume(年月), Number(号), Date(日付), Title(会議録の表題), SpeakerPosition(発言者の役職), SpeakerName(発言者の氏名), QuestionSpeaker(発言に対応する質問者の氏名と役職), Speaker(発言者の氏名と役職), Utterance(発言)
--

表4 都議会だよりのデータ構造

ID(識別番号), MeetingName(会議名), Date(日付), Headlines(質問者の発言全体の要旨), SubTopic(サブトピック), QuestionSpeaker(質問者), QuestionSummary(質問の要約), AnswerSpeaker(答弁者), AnswerSummary(答弁の要約)
--

4.1 データセット

東京都議会の会議録とその人手要約「都議会だよりの」はWebで公開されている。NTCIR-16 QA Lab-PoliInfo-3[15]が提供するJSON形式で構造化した会議録、都議会だよりのデータを使用した。会議録と都議会だよりのデータ構造をそれぞれ表3,4に示す。会議録データは発言(Utterance)が1文が1フィールドで、その他の項目と共に構成されている。都議会だよりのデータは質問要約(QuestionSummary)や答弁要約(AnswerSummary)などの項目が含まれる。テストでは都議会だよりのAnswerSummaryが空欄で、会議録を基に推論を行う。

4.2 学習データ構築

要約モデルを学習するためには答弁セグメント-進捗要約、答弁セグメント-計画要約のソース・ターゲットペアが必要である。ソース側のセグメントは式2のように選択した。

$$\bar{s}_{train} = \arg \max_{s_i \in S} (\beta \text{score}(s_i, a) + (1 - \beta) \text{score}(s_i, t)) \quad (2)$$

答弁要約aとトピックtから、beta = 0.95でBM25によって各セグメントの類似度を計算、最も高いセ

グメントが *strain* に代入される。

ターゲット側のデータ構築に利用した分類モデルは日本語 Wikipedia 学習済み BERT モデルをベースに転移学習を行った。以上より、進捗要約モデル、計画要約モデルのソース、ターゲットを作成し学習を行った。

4.3 実装の詳細

要約モデルは日本語 Wikipedia 学習済み T5 モデルをベースに実装した。前処理として、Juman++[16] による単語セグメンテーションと、SentencePiece[17] によるサブワードセグメンテーションを行なった。

4.4 比較手法

LAST-1 最終文を抽出し要約とする最も単純な手法。ニュース記事要約タスクでは先頭文を抽出する LEAD-1 が一般に用いられるが、会議録の場合、最後に総括されることが多いため LAST-1 で行う。

要約 1 モデル 答弁の種類(進捗、計画)を考慮しない手法。選択したセグメントを要約モデルのソース、答弁要約をターゲットとする。

NTCIR SOTA モデル 同様のタスクに取り組んだ他チームの手法 [15]。セグメント選択時に質問と答弁の 1 対 1 の対応関係を明確にしている。

4.5 評価結果

要約の評価は、NTCIR の評価方法に準拠し、リファレンスとの N-gram の重なりを図る指標である ROUGE を使用した。評価結果を表 5 に示す。F-1 を見ると、答弁の種類を他の考慮するモデルが比較手法のスコア上回っていることが分かる。答弁の「進捗」「計画」に着目した要約の有効性があったと言える。また、進捗/計画要約 2 モデルと進捗/計画要約 1 モデルを比べると進捗・計画要約モデルは分けて実装した方が高いスコアを示した。1 モデルは要約そのものと進捗・計画要約の特徴を混在してしまい、区別して学習できなかつた可能性がある。

表 6 に進捗/計画要約 2 モデルの出力例を示す。進捗要約を緑色、計画要約を青色で表した。フォーマットの観点から見ると、正解要約よりも生成された要約は統一されている。生成要約はすべて進捗要約、計画要約の 2 文だが、正解要約は 3 例が計画要約のみ、2 例が両方となっている。正解要約は人が重要だと思った部分を可能な限り短い要約に収めているため、要約者の主観で内容が変化する場合がある。

表 5 ROUGE 評価結果

手法	Recall	Precision	F-1
LAST-1	0.4177	0.2167	0.2854
要約 1 モデル	0.2917	0.3376	0.3161
進捗/計画要約 2 モデル	0.5243	0.2926	0.3756
進捗/計画要約 1 モデル	0.4090	0.2998	0.3460
NTCIR SOTA モデル	0.3262	0.3012	0.3132

表 6 進捗/計画要約 2 モデルの出力例

生成された要約	正解要約
被害者に寄り添った支援策を総合的に展開。誰もが安心して暮らせる都市東京の実現に努める。	経済的支援策を着実に実施するとともに 3 年度からの次期支援計画の改定を進める。
帰国者・接触者電話相談センターを設置。24 日からコールセンターの回線を増設し、英語・中国語・韓国語による対応や英語・中国語・韓国語による対応など体制を充実。	コールセンター回線増設や聴覚障害の方のファクシミリ受け付け等、体制を充実。
水辺空間を生かした魅力ある都市の顔づくりを進めるため、未来の東京戦略ビジョンに外堀浄化プロジェクトを位置づけた。水源・水量の確保や導水路の整備方法など検討を進め、水の都にふさわしい東京を実現。	導水の水源、水量の確保や導水路の整備方法等を検討し、着実に進める。
都立産業技術研究センターにバイオ技術を活用したヘルスケア産業への支援を開始。専門部署を設置し、更なる発展を図る。	都立産業技術研究センターで支援開始。2 年度は専門部署を設置し取組を更に発展。
感染症対策本部を中心に課題を検証。課題の検証を行い、今後懸念される感染拡大への備えに万全を期す。	課題を検証しながら見直しを図ってきた。今後とも感染拡大の備えに万全を期す。

で、生成要約は統一されたフォーマットで要約が構成されているため、何を採用するのかという主観的判断が減少している。

5 おわりに

本研究では、議会会議録を基に行政に関する質問の回答を出力する質問応答システムを構築した。人手要約の分析により本稿で定義した「進捗」と「計画」が答弁要約の重要な要素であると仮定した。実際に「進捗」と「計画」の側面から要約する提案システムは性能の向上させることを確認した。特定の側面から要約を生成する提案手法は要約システムの頑健性を高めることにも貢献すると考える。

参考文献

- [1] Di Wang and Eric Nyberg. A long short-term memory model for answer sentence selection in question answering. In **Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)**, pp. 707–712, Beijing, China, July 2015. Association for Computational Linguistics.
- [2] Danqi Chen, Adam Fisch, Jason Weston, and Antoine Bordes. Reading Wikipedia to answer open-domain questions. In **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1870–1879, Vancouver, Canada, July 2017. Association for Computational Linguistics.
- [3] Siva Reddy, Danqi Chen, and Christopher D. Manning. CoQA: A conversational question answering challenge. **Transactions of the Association for Computational Linguistics**, Vol. 7, pp. 249–266, 2019.
- [4] Kevin Knight and Daniel Marcu. Statistics-based summarization - step one: Sentence compression. 05 2000.
- [5] Rada Mihalcea and Paul Tarau. TextRank: Bringing order into text. In **Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing**, pp. 404–411, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [6] Makbule Gulcin Ozsoy, Ferda Nur Alpaslan, and Ilyas Cicekli. Text summarization using latent semantic analysis. **Journal of Information Science**, Vol. 37, No. 4, pp. 405–417, 2011.
- [7] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, **Advances in Neural Information Processing Systems**, Vol. 27. Curran Associates, Inc., 2014.
- [8] Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks. In **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 1073–1083, Vancouver, Canada, July 2017. Association for Computational Linguistics.
- [9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, L ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, **Advances in Neural Information Processing Systems**, Vol. 30. Curran Associates, Inc., 2017.
- [10] Jiaao Chen and Diyi Yang. Multi-view sequence-to-sequence models with conversational structure for abstractive dialogue summarization. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 4106–4118, Online, November 2020. Association for Computational Linguistics.
- [11] Ojas Ahuja, Jiacheng Xu, Akshay Gupta, Kevin Horecka, and Greg Durrett. ASPECTNEWS: Aspect-oriented summarization of news documents. In **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 6494–6506, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [12] Katsumi Kanasaki, Jiawei Yong, Shintaro Kawamura, Shoichi Naitoh, and Kiyohiko Shinomiya. **Cue-Phrase-Based Text Segmentation and Optimal Segment Concatenation for the NTCIR-14 QA Lab-PoliInfo Task**, pp. 85–96. 11 2019.
- [13] Giambattista Amati. **BM25**, pp. 257–260. Springer US, Boston, MA, 2009.
- [14] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. **CoRR**, Vol. abs/1908.10084, , 2019.
- [15] Hideyuki Shibuki Yasutomo Kimura. Overview of the ntcir-16 qa lab-poliinfo-3 task. In **Proceedings of The 16th NTCIR Conference**, 2022.
- [16] Arseny Tolmachev, Daisuke Kawahara, and Sadao Kurohashi. Juman++: A morphological analysis toolkit for scriptio continua. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 54–59, Brussels, Belgium, November 2018. Association for Computational Linguistics.
- [17] Taku Kudo and John Richardson. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 66–71, Brussels, Belgium, November 2018. Association for Computational Linguistics.