

地方議会会議録の要約に向けて

○^{1,2} 木村泰知 ² 関根聡 ^{2,3} 乾健太郎

¹ 小樽商科大学 ² 理化学研究所 AIP センター ³ 東北大学

kimura@res.otaru-uc.ac.jp

1 はじめに

近年、国や自治体などが保有する公共データの活用に期待が高まっており、自治体のオープンデータ化が進んでいる。ウェブ上に公開されている公共データの一つに地方議会会議録がある。地方議会会議録は、議会における議論の過程を書き起こしたテキストデータであり、有益な情報が含まれているが、すべての発言が書き起こされており、膨大なテキストである。しかしながら、現在の市民の政治参加や市民活動に地方議会の議論を役立てるために、会議録に対する容易なアクセスができない。容易なアクセスには、柔軟な検索などの技術の開発が行われているが、議会会議録の特殊性に即した形の要約が求められている。ここでいう特殊性とは、議会会議録における固有の議論の構造と挨拶や背景説明を含めた冗長な発言のことである。このような特殊性を鑑みた要約のあり方について実際の議会会議録を分析し、理想的な要約のあり方について議論する。

我々は、以前から、地方議会会議録を対象として、自然言語処理技術を用いた政治課題の抽出や既存の検索エンジンを利用することで情報アクセスの向上を目指してきた [1][2]。しかしながら、政治課題の抽出や情報検索の結果を提示するだけでは、背景や文脈がわからず、内容を把握することが困難であった。

そこで、本研究では、地方議会会議録の特徴について整理するとともに、地方議会会議録に対して、どのような要約をするべきか検討する。本稿では、地方議会会議録の説明および分析の後に、議会会議録の要約のあり方について述べる。

2 地方議会会議録とは

本節では、地方議会会議録の要約に向けて、地方議会会議録の概観および特徴について述べる。

表1に全国の自治体数、地方議会会議録のウェブ公開率¹、議員定数、議員実数(男女別)²をそれぞれ「都道府県」「市区」「町村」別に集計した値を示す。地方議会会議録は、地域住民に対して議会の内容を公開することを目的として作成されており、86%の自治体がウェブ上に公開している。

表 1: 自治体数, 会議録のウェブ公開率, 議員数

	都道府県	市区	町村	合計
自治体数	47	813	928	1,788
会議録のウェブ公開率	100%	100%	73%	86%
議員定数	2,687	19,519	11,230	33,436

地方議会会議録の特徴としては、いつ(発言の日付)、どこで(発言の場所)、だれが(発言者)、なにを発言したのか、明らかにできることが挙げられる。また、議題を議案番号に省略して議論する点や、議員による質問に対して首長が答弁する点も特徴といえる。これらの特徴について、簡単に説明する。

図1に会議録の例を示す。

平成26年 9月 定例会 (第368号)

平成二十六年山形県議会九月定例会会議録 **どこ (自治体・議会種別)**

平成二十六年九月十八日 (木曜日) 午前十一時十一分 開会 いつ (年月日)

議事日程第一号
平成二十六年九月十八日 (木曜日) 午前十時開議

第一 会議録署名議員の指名
第二 会期の決定
第三 議第四百二十二号 平成二十六年度山形県一般会計補正予算 (第三号)

P.4 ○議長 (鈴木正法議員) **だれが(名前・役職)**

発言 ◆質問 ○答弁

○議長 (鈴木正法議員) **ただいまより平成二十六年山形県議会九月定例会を開会いたします。直ちに本日の会議を開きます。なにを (発言内容)**

図 1: 地方議会会議録の公開例

¹ウェブ公開率は地方議会会議録コーパスプロジェクトにより調査した数値である。http://local-politics.jp.

²総務省「地方公共団体の議会の議員及び長の所属党派別人員調査 (平成28年12月31日現在)」

発言の日付は、会議録に必ず記載されており、開催日に加えて、開始時間も記載されていることがある。

発言の場所とは、定期的にかかれる「定例会」や必要に応じて開かれる「臨時会」のような議会の種別のことである。定例会や臨時会では、議員全員で構成される「本会議」や専門の議員で構成される「委員会」が開かれ、議案の審議や審査が行われる³。

地方議会における発言者は、知事、副知事、議員に加えて、知事の代弁者としての職員である。また、議長は、議員の一人が務めることとなり、議員でも、議事進行役になる。そのため、地方議会における発言者の立場は「知事側（副知事・代弁者）」「議長」「議員」の3つに分類できる。

地方議会には、首長の行政執行に対するチェック機能という役割がある。そのため、議員が首長をチェックする意味で質問が行われ、首長（あるいは首長の代理として職員）が質問に対して答弁する形式となっている。このことから、地方議会における発言は、発言者の立場により、おおまかに分類することができ、議員による「質問」、知事による「答弁」、議長による「議事進行」に分類できる。質問と答弁の方式は「一問一答」、質問を一括で読み上げる「一括質問一括答弁」、あるいは、それらを状況に応じて利用する場合がある。

3 地方議会会議録コーパス

我々は、ウェブ上に公開されている地方議会会議録を研究利用できるように収集・整理を進めている。本コーパスは、統計量を扱う分野にも利用できることを考えているため、網羅性および正確性を重視している。そのため、収集対象および収集期間を決めて、全ての会議録を収集することとした。収集対象は全国47都道府県議会の本会議としており、収集期間は平成23年4月の統一地方選挙から平成27年4月の統一地方選挙の前（平成27年3月）までの4年間である。その結果、全国47の都道府県の本会議会議録を全て収集したデータは、約400万文、1.5GBとなった。

収集した会議録データは、統一したデータ構造で管理するために、表2のような17項目からなる「発言テーブル」に格納した。また、議員のプロフィール情報は名前やふりがななどの8項目からなる「議員テーブル」へ格納した。

さらに、本コーパスは、会議録中のすべての発言に対して「発言者」の情報を正確に付与している[3]。例

³自治体によっては「招集会議」「臨時会議」「通常会議」と呼ばれることや定例会の回数が異なることもある。

表 2: 発言テーブル

	フィールド	説明
1	識別子	発言 ID
2	都道府県名	例:「北海道」「青森県」
3	回	会議の回数(第〇回)
4	号	会議の号(〇日)
5	年	開催年(和暦) 23, 24, 25...
6	月	開催月 1~12
7	日	開催日 1~31
8	会議種別	定例会・臨時会
9	表題	例:「第 21 回岩手県議会臨時会」
10	役割	例:「議長」「質問者」「答弁者」
11	発言者 ID	議員テーブルを参照する外部キー
12	発言者名	例:「工藤孝男」「佐々木博」
13	役職	例:「知事」「教育長」
14	発言文	例:「次に、冬の節電対策について...」
15	発言以外	例:「(拍手)」 「[…君登壇]」
16	URL	ダウンロード元の URL
17	保存場所	HTML ファイルの保存場所

えば、「梅沢佳一」や「梅澤よしかず」などと記載されている場合でも同一人物として識別できる。収集および整理の結果、本コーパスは、異なり発言者数 5,942 人、総文字数 24,784,262 文字、総行数 4,395,876 行⁴、1 行あたりの平均文字数 56.4 文字であることを確認した。

4 要約に向けて

本節では、地方議会会議録の要約に向けた分析の結果を述べるとともに、実現方法について考える。

4.1 会議録の問題点

図2に東京都議会の平成23年本会議の会議録の例を示す。図2では、山下太郎議員の質問とその質問に対する知事の答弁を簡単にまとめたものである。都議会の本会議は、一括質問一括答弁であり、質問を一括で読み上げているため、山下太郎議員の1回の質問だけで269行を超えており、一回の発言が19,204文字となっており、21の質問が含まれている。破線で囲まれた議員の「東日本大震災」に関連箇所だけでも26行(1,609文字)である。また、この質問に対する知事の答弁は58行(4,329文字)となっており、そのなかで、「被災地支援」についての答弁が6行(450文字)である。さらに、知事は、一回の発言で、議員の21の質問に全て回答しているわけではなく、他の質問について、担当職員から回答することを最後に述べているように、答弁を続けている。

この例からわかるように、3つの問題がある。1点目は、東京都議会の本会議が、一括質問一括答弁の形

⁴1行の区切りは、句点「。」および「改行コード」としている。

東京都議会会議録（地方議会会議録コーパス）		
議員 山下太郎 269行 (19,204文字)	256行目	私は、都議会民主党を代表して、都政の主要課題について知事並びに関係局長に伺います。
	266行目	...
	267行目	まず、東日本大震災における被災地支援と東京の防災対策について伺います。
	268行目	三月十一日、マグニチュード九・〇、最高震度七の強く長い揺れが東日本一帯を襲うとともに、大津波、海砂を巻き込んだ黒く重い海水の塊が太平洋沿岸の防波堤を軒並み破壊し、海水や瓦れきが市街地に流れ込み、甚大な被害を引き起こしました。
	269行目	福島第一原子力発電所にも大津波が押し寄せ、冷却電源を失った原子炉建屋は爆発、格納容器が損傷して、放射性物質が広範囲に拡散しました。
	276行目	原発周辺の住民の皆さんは、自宅があるのに帰れない深刻な状況が続いています。
知事 石原慎太郎 58行 (4,329文字)	276行目	被災地のニーズを的確に把握し、被災地、被災者が真に必要とする支援に今後とも継続して取り組むべきと考えますが、知事の見解を伺います。現在、都内には福島県などから自主避難してきた約五千名の避難者の皆さんが都営住宅などに仮住まいをいらっっしゃいます。
	292行目	次に、原子力発電所の安全性について伺います。
	524行目	ご清聴ありがとうございました。
	527行目	山下太郎議員の代表質問にお答えいたします。
528行目	まず、被災地支援についてであります。震災発生から三カ月たちましたが、...	
529行目	電力、農林水産物等の供給の多くを被災地に依存してきた首都東京が、全国の先頭に立って刻苦する被災地を支援していくことは当然であります。	
532行目	国の対応は、厳しい被災地の現状に対して後手後手に回っておりますが、東京は、今後とも復興に向けみずから踏み出す被災地、被災者を強力に後押ししていきたいと思っております。	
584行目	他の質問については、教育長、東京都技監及び関係局長から答弁します。	

図 2: 東京都議会会議録の例

式をとっていることから、発言者の一回の発言が長いことである。質問および答弁が長いこと、破線で囲んだように、質問内容ごとに区切る必要がある。2点目は、発言には、背景の説明や挨拶などの質問以外の内容が含まれることであり、重要な箇所を見つけづらいことである。例えば、図2の部分では背景の説明が3行以上述べられている。3点目は、議員の質問に対する知事の答弁が離れた位置にあり、関係性をみつけるのが困難なことである。

これらの地方議会特有の問題を考慮しつつ、議論をおおまかに理解できる要約が望まれる。

4.2 議会だよりの存在

どのような要約がよいか調査していたところ「議会だより」が参考なることがわかった。「議会だより」は、地方議会の内容を自治体がまとめて住民に提供している情報誌であり、会議録に対して人手で作成した要約文章といえる。表3は、47都道府県の自治体が「議会だより」を提供しているのか、また、どのように公開したものである。ほとんどの自治体の「議会だより」には、議員の質問に対する知事の答弁の要約が含まれていた。これらのことから、地方議会会議録の要約には、質問に対する答弁の対応関係が必要であることを確認した。

我々は、議会だよりが、会議録における要約の理想的な要約と考えている。全国の自治体では、議会だよりを公開していない自治体もあることから、自動化の

必要性もある。

表 3: 「議会だより」の公開の有無および公開方法

公開あり			公開なし	合計
HTML+PDF	HTML	PDF	9	
10	2	26	9	47

4.3 会議録と議会だよりの関係の分析

本節では、東京都の「議会だより」がどのように要約しているのかを調査する。また、会議録と議会だよりの関係の分析を行う。

図3に平成23年第2回定例会の「都議会だより293号」の例を示す。図3では、議員が「東日本大震災」について質問しており、質問内容が「被災地が真に必要なとする支援に継続して取り組むべき。知事の見解は。」と要約されている。また、知事の答弁は「全国の先頭に立ち刻苦する被災地を支援するのは当然。今後も強力に後押しする。」と要約されている。

ここで、図3の質問と答弁がの要約が「地方議会会議録コーパス」において、どのように記載されているか図2と比較する。図3の発言者（質問者と答弁者）、トピック、内容の要約に、それぞれ、青、緑、赤の色分けをして記述している。この結果から、我々は、議会会議録において、発言者、トピック、質問に対する

⁴<http://www.gikai.metro.tokyo.jp/newsletter/293/03-1.html>

簡単な答弁があることが、理想的な要約であると考えている。

■ 発言者 (質問者・答弁者) ■ トピック ■ 内容の要約	山下太郎 (民主党)
東日本大震災	被災地が真に必要なとする支援に継続して取り組むべき。知事の見解は。 知事 全国の前頭に立ち刻する被災地を支援するのは当然。今後も強力に後押しする。
東京の総合防災力	更に高める取組が必要。 知事 新視点の対応加えた防災対応指針を11月に策定し防災力向上の道筋を早期に示す。
原子力発電所	安全性に対する基本認識は。 知事 原子力利用に当たって安全の確保は当然。信頼性失ってきた姿勢を反省し再考すべき。同時にエネルギー確保は国家・社会の存立に直結、戦略を速やかに構え直す必要あり。

図 3: 都議会だより 293 号の例

4.4 会議録要約の実現に向けて

我々は、「議会だより」「地方議会議録コーパス」の2つのリソースを利用した要約技術の要素開発を行う。いままでの分析結果を踏まえて、技術開発の提案を行う。具体的には、下記に示した4つの処理により質問と答弁の関係をまとめることが、会議録の要約といえる。

1. 発言の分割

発言の分割は、「まず」「次に」のような接続詞を手がかりにすることができる。また、分散表現を利用した文章分類の問題にする方法がある。

2. 分割結果へのラベル付与

発言のラベルは、議員、知事、議長によって異なる。議員の発言には、質問、挨拶、背景説明などがある。知事の発言には、答弁、挨拶、所信表明などがある。議長の発言には、議事進行、報告、紹介、採決などがある。ラベル付与に対しても、手がかり表現が有効と考えられる。例えば、質問は「伺います」という文末表現になることが多い。このようなタイプの分類は、重要箇所抽出に役立つ。

3. 分割結果内の重要箇所抽出

会議録の要約では、発言内容を変更せずに、重

要箇所を抽出する必要ある分割結果内の重要箇所は、TFIDF などを利用できる。

4. 質問と答弁の関係の認識

質問と答弁の認識は、議員の質問と知事の答弁を結びつけることであり、発言内容の類似度を利用できる。

地方議会議録の要約は、議会だより、および、地方議会議録コーパスの対応関係を用いることで、新しい要約技術の開発ができる。要約のトレーニングデータが大量にでき、それに基づいた要約技術を機械学習などを利用して、構築していきたい。

5 おわりに

本稿では、地方議会議録の特徴について整理するとともに、地方議会議録に対して、どのような要約をするべきか、議会だよりの要約を参考にしながら、調査を行うとともに、議会議録の要約のあり方について述べた。

今後、我々は、地方議会議録コーパスを研究者に向けて公開する予定であり、数多くの研究者と協力しながら、研究を進めることを考えている。

謝辞

本研究は JSPS 科研費 JP16H02912 の助成を受けたものです。

参考文献

- [1] 木村泰知, 関根聡, 主辞に基づく政治問題抽出手法, 人工知能学会論文誌, Vol.28, No.4, pp.370-378, 2013.
- [2] Yasutomo Kimura et al., Creating Japanese Political Corpus from Local Assembly Minutes of 47 Prefectures, Coling 2016, The 12th Workshop on Asian Language Resources, pp.78-85, 2016.
- [3] 木村泰知, 内田ゆず, 高丸圭一, 都道府県議会議録のパネルデータ作成に向けた発言者情報の付与, 第 33 回ファジィシステムシンポジウム講演論文集, pp.701-706, 2017.