

# 国会および地方議会の予算審議を対象とした Budget Argument Mining タスクの提案

永淵景祐  
小樽商科大学

g2018268@edu.otaru-uc.jp

土屋彩夏  
小樽商科大学

木村泰知  
小樽商科大学

kimura@res.otaru-uc.jp

## 1 はじめに

政治には、収入と支出を考慮し、お金の使い道を定める予算作成の役割がある。国の予算は、内閣で予算案が作成され、その予算案をもとに国会で議論された後に成立する。また、地方自治体の予算は、知事や市長により予算案が作成され、議会で審議された後に成立する。このような過程を経て成立する予算は、どのような背景に基づいて予算案が作成され、どのような議論を経て成立しているのかを把握しづらい。

従来から、政治学や経済学の分野において、国や地方自治体の予算に関する研究が行われている。予算過程では、(1)財政収支計画の作成、(2)審議、(3)執行、(4)決算、の順番に進み、租税の配分や経費の配分について意思決定が行われる [1]。予算を含む審議の分析は、国会、あるいは、地方議会の会議録を対象として、TFIDFを用いた分析、あるいは、共起ネットワークを用いた分析を行っている [2][3]。また、NTCIR-15 QA Lab-PoliInfo-2<sup>1)</sup>では、東京都議会会議録を対象として、議論構造を考慮した自動要約、議案に対する各会派の賛否分類、法律名の表記揺れおよび曖昧性 Entity Linking などの問題を解決する Shared task を開催している [4]。しかしながら、議会における予算審議の過程をわかりやすく自動で提示する方法については、研究の余地がある。

そこで、我々は、国や自治体の予算成立までの議論に着目し、どのような議論に基づいて成立したのかを簡単に把握できるシステムを開発することを目標とする。本研究では、国会や地方議会における予算案の議論を対象として、概算要求の主要事項一覧と予算に関する議論を結びつけるタスクを提案する<sup>2)</sup>。国会や地方議会では、予算に関する議論が

「議会会議録」として記録されている。そこで、本研究では、国会および地方議会における概算要求に記載されている「主要事項」および「金額表現」に基づいて対応する議論を議会会議録から見つけ、結びつけるタスクを提案する。

本研究の貢献は、下記の3つである。

1. 予算項目と議論を結びつけるタスクの提案
2. 議会の発言に含まれる金額表現の調査
3. 予算審議に適した議論構造のラベルの提案

## 2 関連研究

本研究では、予算審議の議論における金額表現に焦点を当てる。テキストに含まれる金額表現の推定は、固有表現抽出における「金額」の抽出タスクとみなすことができる [5, 6, 7]。金額表現は、人名や組織名と比較すると、高い精度で推定可能な固有表現である。

金額表現に関する研究には、金融関連のテキストを対象とした研究がある。例えば、NTCIR15のFinuNumでは、金融関連のツイートに含まれる数値表現が対象となる項目に対して、関連しているのか、関連していないのか、2値分類を行うタスクである<sup>3)</sup>。ここでは、ツイートを対象としており、関連性の有無だけを扱っており、議論の構造を対象としていない。

数字を含む自然言語文を理解する研究では、数量的常識の獲得を目指す研究があり、文章中の物体や現象の長さや重さなど、実世界における数量的特徴に関する知識を用いて、数字の予測に関する研究を行っている [8][9]。

国会会議録は、検索APIが公開されていることから、議員の発言を対象としたテキスト分析が行われている [10]。

1) <https://poliinfo2.net/>

2) 例えば、令和3年度の厚生労働省予算概算要求の主要事項一覧は、<https://www.mhlw.go.jp/wp/yosan/yosan/21syokan/>

のサイトで閲覧できる。

3) <https://sites.google.com/nlg.csie.ntu.edu.tw/finnum2020/finnum-2>

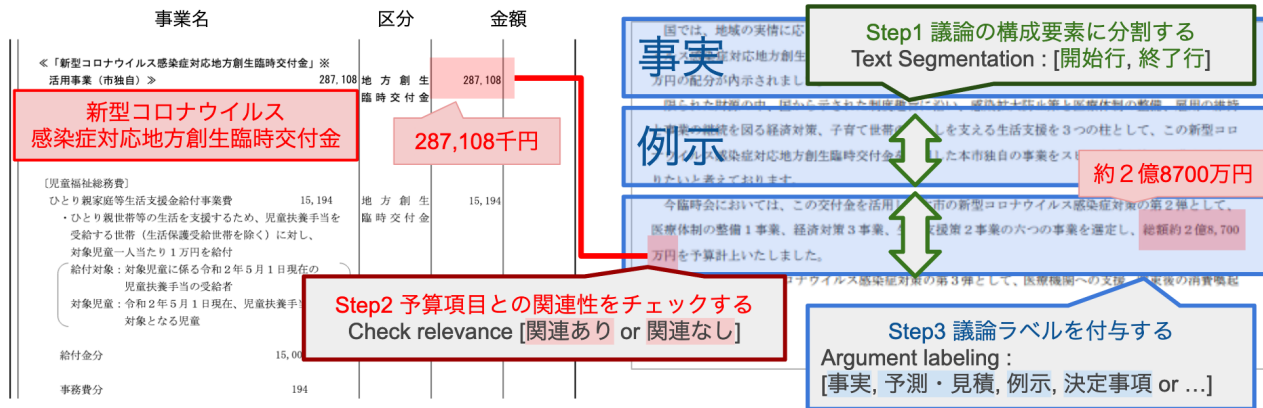


図1 小樽市の予算総括表と会議録を用いたタスクの概要

Argument mining は、自然言語処理のアプローチにより議論構造をとらえる研究を行っている [11]. Argument 分析の手順は、一般的に、Text segmentation, Classification(Argument or Non-argument), Simple Structure, Refined Structure の順番に行く。

### 3 Budget Argument Mining

本稿では、国会、および、地方議会における「予算審議に関する項目」と「議会会議録に含まれる関連した議論」を結びつけるタスクを **Budget Argument Mining** と呼ぶ。図2には、小樽市の「予算総括表」と「議会会議録」を例として、**Budget Argument Mining** タスクのイメージを示す。本タスクは、既存の予算事項一覧を起点として、予算が決まる過程である議論、つまり、会議録に含まれる関連した議員の発言に結びつける。

具体的には、次の3つのステップにより、予算項目と関連する議論を結びつける。

1. Step 1 **Text segmentation** : 議会会議録に含まれる発言を議論の構成要素に分割する。
2. Step 2 **Check relevance** : 予算項目との関連性を「関連あり」「関連なし」の2値で推定する。
3. Step 3 **Argument labeling** : 議論の構成要素に対して、予算審議の議論ラベルを付与する。

図2の予算総括表に「新型コロナウイルス感染症対応地方創生臨時交付金」の事業名があり、その金額は287,108千円と記載されている。ここで記載されている287,108千円の議論は、小樽市議会会議録の第1回臨時会で議論されている。小樽市議会の会議録の場合、市長や議員の発言内容ごとに段落単位で記述されているため、**Text segmentation** は、既に

が行われている状態である。**Check relevance** は、図2にある議論の構成要素の発言（以下「発言」と呼ぶ）に対して、新型コロナウイルスに対する関連の有無を推定する。**Argument labeling** は、関連があると推定された発言に対して「事実」「予測・見積」などの予算審議の議論ラベルを付与する。

### 4 金額表現と議論ラベル

#### 4.1 予算審議に関する注釈付け

**金額表現** : 予算の金額表現には、図2の予算総括表に記載されている金額（287,108千円）に加えて、議会会議録に記載されている金額（約2億8700万円）がある。特に、議会会議録に含まれる金額表現は、表記揺れ、曖昧性の問題がある。そこで、本節では、予算審議における金額表現がどのように記載されているのか、議会会議録を対象に明らかにする。

**議論ラベル** : 予算審議の議論ラベルとは、議会会議録に含まれる議員の発言に対して、議論の流れを理解するために必要となるラベルである。そこで、本節では、予算審議における会議録に対して、議論ラベルの注釈付けを行いながら、適切な議論ラベルについて検討する。

#### 4.2 注釈付けの手順

注釈付けは、下記の手順で進める。

1. 国会、地方議会を対象に金額表現、および、議論ラベルを作成する。
2. 適切な議論ラベルであるか確認するために、新たな議会会議録を対象として、注釈付けを行う

ことで確認する。

まず、議論ラベルを作成するために、国会会議録、北海道議会会議録、小樽市議会会議録を対象として、金額表現を含む発言の調査を行う。そのなかでも「新型コロナウイルス」に関連した議論に焦点を当て、4月下旬から5月上旬の議会を選択する。

次に、適切な議論ラベルであるかを確認するために、国会会議録の予算審議を対象として、注釈付けを行う。対象とする委員会は、令和3年度の厚生労働省の概算要求に関する議案であり、新型コロナウイルスの内容も含まれる。

### 4.3 金額表現を含む発言の調査

我々は、国会衆議院予算委員会、北海道議会、小樽市議の議会録を対象として、金額表現を含む発言について調査した。表1に金額表現を含む発言の調査結果を示す。表1から国会は金額表現を含む発言の割合が多いことがわかる。また、国会は金額表現の表記揺れはなく、全て漢数字（例えば、二・二兆円）で記述されている。一方、北海道議会、および、小樽市議会は、アラビア数字で記述されていた。

表1 金額表現を含む発言の調査結果

議会 (日付)	金額表現 件数 (A)	全文字数 (B)	割合 A / B
第201回国会衆議院 予算委員会 (令和2年4月28日)	150	123,706	0.0012
北海道議会第1回臨時会 (令和2年4月28日)	24	53,668	0.0004
小樽市議会第1回臨時会 (令和2年5月22日)	14	17,761	0.0008

また、金額表現を含む発言に付与する議論ラベルを作成した。

**決定事項：** 国・自治体の政策などで決定したこと、あるいは、それに関わる金額。特に、その年の予算委員会で決められた予算など

**例示：** 過去の事例で用いられた金額

**事実：** 現在の議論や政策に関連のない単なる事実としての金額。例、前年度等の予算や金額

**予測・見積：** 見積額や予測金額など

**訂正事項：** 誤りについて訂正された金額

**メディア：** メディアで報じられた金額

**提案：** 決定に至っていない提案事項。例、野党からの改善案に含まれる金額

**目安：** おおよその基準となる金額

**質問・確認事項：** 質問や確認に含まれる金額

### 4.4 議論ラベルの付与

表2に4.3の議案ラベルを用いて、第201回国会衆議院予算委員会第26号（令和2年6月9日）の会議録に対して、議論ラベルを注釈付けした結果を示す。表2からわかるように、決定事項、事実、例示に関する発言が多いことがわかる。金額表現の合計は182件であり、それらの金額表現は、94発言の中に含まれていた。

表2 金額表現を含む発言に付与された議案ラベル

議論ラベル	金額表現
決定事項	96
例示	18
事実	56
予測・見積	5
訂正事項	3
メディア	1
提案	1
目安	2
質問・確認事項	1
合計	182

さらに、金額表現が含まれる94発言を対象として、厚生労働省の概算要求および補正予算などの項目<sup>4)</sup>を議会の発言と結びつけることが可能であるかを調査した。その結果、6発言が「令和2年度厚生労働省第二次補正予算の項目」と結びつけられることを確認した。下記に、「主要項目」および「金額表現」に基づいて議会会議録の発言と結びつけられた例を示す。

第二次補正予算の項目に結びつけた発言の例

雇用調整助成金は、日額上限八千三百三十円が一万五千円に引き上げられました。

## 5 データフォーマット

本節では、国会会議録を対象として、予算項目と議論を結びつける **Budget Argument Mining** タスクの具体的なデータフォーマットについて述べる。

### 5.1 国会会議録について

国会会議は、本会議、常任委員会、特別委員会、憲法審査会、情報監視審査会、政治倫理審査会の5つに分けることができる<sup>5)</sup>。現在検討している対象

4) <https://www.mhlw.go.jp/wp/yosan/yosan/>

5) 政治倫理審査会の会議録は公開されていない。

(単位：百万円)

項目	主要事項	令和2年度 予算額	令和3年度 要求・要望額
第1 新型コロナウイルス感染症から国民のいのちを守るための体制確保	1 新型コロナウイルスと戦う医療・福祉提供体制の確保	7,693	690 + <span style="border: 1px solid black; padding: 0 2px;">緊要</span>
	2 検査体制の充実、ワクチン・治療薬の開発・確保	0	100 + <span style="border: 1px solid black; padding: 0 2px;">緊要</span>
	3 感染拡大防止に向けた研究開発の推進	1,303	3,276 + <span style="border: 1px solid black; padding: 0 2px;">緊要</span>

図2 国会、令和3年度厚生労働省予算概算要求の主要事項一覧表

会議録は、衆議院と参議院の本会議と委員会を合わせた46の会議録である。

## 5.2 国会会議録 API について

国会の会議録については、国会会議録 API を用いて、収集する。国会会議録検索システムでは、登録されているデータを検索し、取得するための外部提供インターフェイス (API: Application Programming Interface) を提供している<sup>6)</sup>。検索用 API を利用することで、検索リクエストに対し、XML 形式又は JSON 形式でデータを取得することができる。検索用 API は、(1) 会議単位簡易出力、(2) 会議単位出力、(3) 発言単位出力 の3種類がある。本研究では、議長や議員の発言を全て取得する必要があるため、(2) 会議単位出力を用いる。

本タスクデータセットには、国会会議録を対象とする場合、国会会議録 API の情報に、下記の項目 `checkRelevance`, `relatedBudgetID`, `argumentClass`, `moneyExpression` を追加する。

## 5.3 Json のフォーマット

本節では、本タスクの Json フォーマットについて述べる。入力対象データは、「令和3年度厚生労働省予算概算要求の主要事項一覧表」である。検索対象データは、国会衆議院第201回予算委員会26号(2020年6月9日)である。

図2の令和3年度厚生労働省予算概算要求の主要事項一覧表に対応する Json フォーマット、および、国会会議録の Json フォーマットを下記に示す。

Listing 1 予算一覧表の Json フォーマット

```

{
  "budgetID": 識別子,
  "detail": [
    {
      "bill": 令和2年度厚生労働省第二次補正予算,
      "item": 項目,
      "mainTopic": 主要事項,
      "subTopic": 事項,
      "budget": 令和2年度予算額,
      "request": 令和3年度予算額要求・要望額,
      "urgent": 緊要
    }
  ],
}

```

6) <https://kokkai.ndl.go.jp/api.html>

Listing 2 国会会議録の Json フォーマット

```

{
  "issueID": 会議録ID,
  "imageKind": イメージ種別,
  "searchObject": 検索対象箇所(議事冒頭・本文),
  "session": 国会回次,
  "nameOfHouse": 院名,
  "nameOfMeeting": 会議名,
  "issue": 号数,
  "date": 開催日付,
  "closing": 閉会中フラグ,
  "speechRecord": [
    {
      "speechID": 発言ID,
      "speechOrder": 発言番号,
      "speaker": 発言者名,
      "speakerYomi": 発言者よみ,
      "speakerGroup": 発言者所属党派,
      "speakerPosition": 発言者肩書き,
      "speakerRole": 発言者役割,
      "speech": 発言,
      "startPage": 発言が掲載されている開始ページ,
      "createTime": レコード登録日,
      "updateTime": レコード更新日時,
      "speechURL": 発言URL,
      "checkRelevance": null,
      "moneyExpression": [],
      "relatedBudgetID": [],
      "argumentClass": []
    }
  ],
  "meetingURL": 会議録テキスト表示画面のURL,
  "pdfURL": 会議録表示画面の※PDFURL,
}

```

## 6 おわりに

本稿では、国会および地方議会における予算審議に関する「主要事項」および「金額表現」に基づいて関連する発言を結びつけるタスクについて提案した。また、タスク設計に向けて、国会、および、小樽市議会の予算項目を分析することにより、議員の発言に含まれる金額表現の特徴を述べるとともに、予算審議に適した議論構造のラベルを提案した。

今後は、**Budget Argument Mining** タスクを Shared task として開催する予定である。

## 謝辞

本研究は JSPS 科研費 18K00632, 20K00576, 20H00059, 20K01736 およびセコム科学技術振興財団の助成を受けたものである。

## 参考文献

- [1] 横田茂. [研究ノート] 日本の予算制度と予算過程 : その特質の形成. 關西大學商學論集, Vol. 64, No. 1, pp. 77–103, jun 2019.
- [2] 名取良太, 田中智和, 岡本哲和, 石橋章市朗, 梶原晶, 坂本治也, 秦正樹. 地方議会の審議過程 : テキスト分析による定量化の試み. mar 2020.
- [3] 増田正. 計量テキスト分析によるわが国地方議会の審議内容を可視化する方法について. 地域政策研究, Vol. 19, No. 3, pp. 161–175, feb 2017.
- [4] Yasutomo Kimura, Hideyuki Shibuki, Hokuto Ototake, Yuzu Uchida, Keiichi Takamaru, Madoka Ishioroshi, Teruko Mitamura, Masaharu Yoshioka, Tomoyoshi Akiba, Yasuhiro Ogawa, Minoru Sasaki, Kenichi Yokote, Tatsunori Mori, Kenji Araki, Satoshi Sekine, and Noriko Kando. Overview of the ntcir-15 qa lab-poliinfo task. *Proceedings of The 15th NTCIR Conference*, 12 2020.
- [5] 関根聡, 井佐原均. Irex : 情報検索、情報抽出コンテスト. Technical Report 81(1998-FI-051), ニューヨーク大学コンピュータサイエンス学科, 郵政省通信総合研究所, sep 1998.
- [6] 山田寛康, 工藤拓, 松本裕治. Support vector machine を用いた日本語固有表現抽出. 情報処理学会論文誌, Vol. 43, No. 1, pp. 44–53, jan 2002.
- [7] Satoshi Sekine, Kiyoshi Sudo, and Chikashi Nobata. Extended named entity hierarchy. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas, Canary Islands - Spain, May 2002. European Language Resources Association (ELRA).
- [8] Georgios Spithourakis and Sebastian Riedel. Numeracy for language models: Evaluating and improving their ability to predict numbers. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2104–2115, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [9] 阪本拓功, 相澤彰子. 言語モデルを用いたテキスト中の数字の予測. 情報処理学会 自然言語処理研究会, 2020.
- [10] AkitakaMatsuo and Fukumoto Kentaro. Legislators' sentiment analysis supervised by legislators. Technical Report 38, Department of Government, University of Essex, Department of Political Science, Gakushuin University, nov 2020.
- [11] John Lawrence and Chris Reed. Argument mining: A survey. *Computational Linguistics*, Vol. 45, No. 4, pp. 765–818, December 2019.