

国会議事録を使用した 政党ごとのスタンス変遷の分析

尾崎 慎太郎 横山 大作
明治大学 理工学部 情報科学科
{ee207060, dyokoyama}@meiji.ac.jp

概要

昨今、若者の政治に対する参加意欲の低下が日本では深刻化している。その原因には、様々な理由があるが実際に実現されようとしているのかがわからないで似たような発言をしているなどの声が多い。本研究では、大量の国会議事録から政党ごとの政策に対するスタンスの変遷を分析し、一貫性があるかどうかを確認できることを目指した。そのため、国会での発言から政党のスタンスを分類、また発言を要約することで議案を特定し、マニフェストに掲載された5つの代表的な政策について分析した。その結果、一貫性が見られた政党を確認できた一方で、選挙公約で話していたことを実現していないと思われる政党も確認した。

1 はじめに

若者の政治への参加率が低いことが日本の社会問題として挙げられている。例えば、図1に示した、総務省が公開した「国政選挙における年代別投票率について」[1]によると、令和3年10月に行われた第49回衆議院議員総選挙に関して、全年代を通じた投票率が55.93%であるのに対し、10代が43.21%、20代が36.50%となっており、いわゆる若者が政治に参加していないことが目に見えてわかる。

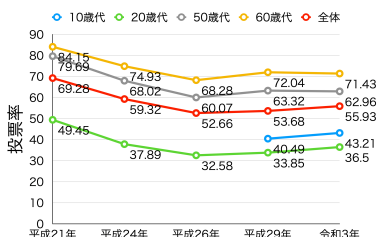


図1 国政選挙における年代別投票率

NHKの調査[2]によると、若者が投票しない理由として「自分たちが望む政策を掲げる政党に投票したものの、実際に実現されようとしているのかがわ

からない」などが挙げられた。この問題については、国会議事録がWeb上でPDFとして公開されており、それを読めば政策の議論について理解することができる。しかし、議事録は発言したものを一言一句記録しており、また難しい専門用語などを使っているため全て読み通すことには困難が伴う。実際に中身を見てみると、2022年の通常国会の中で最も発言が多かった2月16日の議事録には約130万文字の発言があった。そこで、公開されている議事録の情報を用いて、政党の政策への取り組みを分析できるようなシステムを構築しようと考えた。

具体的に、国会で発言した内容について反対なのかそうでないのかに分類し、またその内容を要約し答弁のテーマを取り出す。そして、まとめた文章とそれに合う政策を取得し比較することで、その政策を実現しようとするスタンスの一貫性を計った。実験を行った結果、ある政党は一貫性が見られた一方で、一貫性が見られないように思われる政党があることも確認できた。

2 関連研究

猫本ら[3]は、BERTを使用した分類器を構築し、東京都議会の議会議事録に対してスタンス分類を行っており、賛否が不明確な討論文から議案に対する賛否を分類している。本研究では、猫本らが問題として挙げていたデータ数が非常に少ないという問題を、国会議事録を使用することによって解消する。

また大杉ら[4]は議題生成に取り組んでいた。ここでも、東京都議会の議事録を使用して構造解析した質問と答弁のペアからSeq2Seq[5]モデルであるTransformer[6]を使用して、議題文を生成した。

他にも、議題の作成という観点では、キーワード抽出や生成の研究がある。抽出したフレーズ候補を似た者同士でクラスタリングし、そのフレーズ候補集団一つ一つをトピックとし、トピックレベル

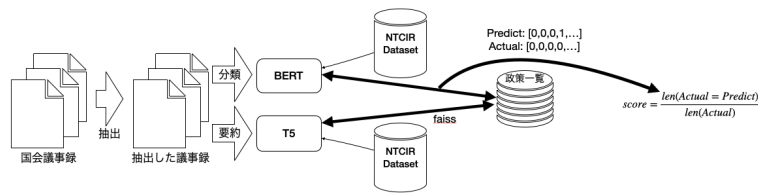


図2 モデルの全体図

で重要度を評価することで、重要なトピックからキーワードを一つずつ抽出することを可能にした TopicRank[7] や、論文のタイトルと概要をペアで入力とし、キーワードを推論させる RNN を使用したキーワード生成 [8] がある。本研究では、要約タスクとしてキーワードを見つけているが、今後の展望としてこれらの手法を用いることも考えられる。

3 提案手法

まず大量の国会議事録に対して発言から、その発言が議題に対して反対なのかそうでないのか分類する。並行して、その発言を要約することで何について話しているかまとめる。そして、まとめた文章とそれに合う政策を比較し、分類したスタンスと選挙公約で話していた政策のスタンスを比較することで分析する。本研究で実装した全体図は図 2 の通りであり、大きく賛否の分類、テーマ要約、類似政策の検索の 3 つのタスクからなる。

賛否の分類 議事録から発言者ごとにまとめた発言を対象に、政策への賛否のスタンスを分類する。分類モデルは東北大学が公開している BERT[9] を使用し、Fine-Tuning では学習データとして NTCIR-14 の QA Lab PoliInfo[10] で用いられたデータセットを使用した。詳しくは、3.2 節で述べる。

テーマ要約 同じく発言者ごとにまとめた発言を対象に、議論の対象となっている政策をよく表すような 1 つの要約文章を生成する。事前学習済みモデルとして、タイトル生成を得意とする T5 モデル [11] を使い、同じく学習データとして NTCIR-14 のデータセットを用いた。詳しくは、3.3 節で述べる。

検索タスク 集計したい対象の政策セットを用意する。議事録の要約した文章ごとに、議論のテーマが最も近い政策を求め、分析する。要約した文章と政策の比較には、Fundamental AI Research が公開している faiss[12] を使用した。選挙ドットコム [13, 14] に記載されているマニフェストを参照するデータベースとして用意し、要約した文章を入力とする。詳しくは、3.4 節で述べる。

表 1 国会理事録 API で取得できるデータ

変数名	データの一例
NameOfHouse	衆議院
NameOfMeeting	予算委員会
Date	2022-01-28
SpeakerGroup	自由民主党
Speech	先ほどの法律に反対します

3.1 国会議事録データの準備

Web 上からの取得 国立国会図書館のホームページ上で、国会で発言されたものを議事録として保存されており、API¹⁾を通じて発言データを取得することができる。取得したデータは下記の表 1 の通りである。この他にも、発言者の名前も取得することができるが、本研究では使用しない。また、本研究において使用したデータは 2021 年と 2022 年の発言を基に分析を行う。

国会議事録のクリーニング処理 API を通じて取得したデータには国会で発言したもの全てが含まれているためノイズが多いため非常に多い。また、議題として上がっている話題とは異なった、関係のない話であったり、議題の方向性とずれた話があり、それらを除く必要がある。本研究ではクリーニング処理の手段として、フィルターの削除、定型文の削除、そして議題とは関係のない不必要な文章の削除を施した。

まず、フィルターに関して、国会で頻繁に発言されることをデータを通じて確認することができた。しかし、これらは不必要であるため削除する。そして次に定型文の削除に関して、国会で質問に返答する際に決まった定型文や、発言をする際に「{ 政党 } の { 発言者名 } です。」と申し上げることが慣例として挙げられる。これらの情報はメタデータに含まれているため必要がないと判断し、{ {1,10}(党 | の会 | 組 | 党. 無所属 | の会. 無所属 | 無所属クラブ | 政務官 | 大臣 | 新緑風会)(の |)(.*) (です | でございます | と申します) } のような正規表現によって政党と発言者名を削

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

除した。また、決まった定型文の一例として、「ご質問ありがとうございます」、「御答弁ありがとうございます」、「この際、一言ご挨拶申し上げます」などがあり、これらも削除した。付録の図 4 に他の例も記載している。

他にも、国会で発言をしている政党にはさまざまな種類がある。例えば、組み合わせが複雑なものとして「立憲民主党・無所属」や「国民民主党・新緑風会」、「公明党・国民の声」、「国民民主党・無所属クラブ」などがある。本研究では議席数の多い6党（「自由民主党」、「公明党」、「立憲民主党」、「日本共産党」、「日本維新の会」、「社会民主党」）に絞って研究を行った。また、国会議事録には具体的なテーマに触れず抽象的な発言が多くある。本研究では、明確に賛否がわかる文章のみを抽出することを目指し、具体的に正規表現で「努めてまいります」、「承知しております」、「積極的にやりたい」、「すべきだと思うんです」、「お聞きしたいというふうに思います」などのような文章が含まれているものを分類並びにテーマ要約を行うための入力文章として抽出した。最後に、抽出前後の国会議事録の発言数を付録の表 6 に記載する。

3.2 国会議事録の分類

Fine-Tuning には NTCIR-14 QA Lab PoliInfo で使用された公開されているデータセットをモデルの学習データとして使用した。中身は東京都議会の議事録であり、発言とそのスタンスがペアとして 31,807 件あった。データの中で発言とそのスタンスを使用する。Classification Task では賛成と反対、そしてそのどちらでもない(その他)の3つのクラスに分類されたデータであったが、本研究では反対か反対でないかの2分類にするために、賛成とその他を1つのクラスとした。背景として、政策に対し中立の立場を取ることができてしまうと、政党ごとに区別がつかなくなってしまうと考えたためである。このことにより、本研究では反対を0、反対ではないを1として研究を進める。学習データ 31,807 件の中身を見てみると、スタンスが1のものが 29,304 件、0のものが 2,503 件であり、不均衡のデータであったため、アンダーサンプリングを実行し、どちらも 2,503 件に絞って実験した。事前学習済みモデルとして東北大学が公開した BERT[9] を使用した。ここに、国会議事録をクリーニング処理したデータの発言を入力とし、その発言から反対かそうでないかを推論させる。

3.3 国会議事録の要約

要約では Fine-Tuning に NTCIR-14 QA Lab PoliInfo で使用された公開されている Summarization Task のデータセットを学習データとして使用した。こちらも中身は東京都議会の議事録であり、発言とその要約文のペアが 742 件あった。また、事前学習済みモデルとして T5[11] を使用した。ここに、クリーニング処理した発言を入力として、そこから要約したものを出力として得る。

3.4 政策と要約文のマッチング

Fundamental AI Research(FAIR) が公開している faiss[12, 15] を用いて実装した。ベクトル化したデータストアには、選挙ドットコム²⁾の政策一覧を用いた refsec:summarization 節で発言を要約したものを入力として近似最近傍探索を実行し、最も近い政策を得る。なお、データストアとする政策に関しては 2021 年 [13] は 6、2022 年 [14] は 20 の政策を用意する。

各党の政策とそのスタンス 正解データとして、選挙ドットコム [13, 14] の各党の政策一覧を用いることとする。具体的に選挙ドットコムには政策一覧とそれぞれの政党ごとに賛成なのか反対なのか、また中立なのかが明記されている。今回は、この政策に書かれてある政策一覧を正解データと仮定して研究を進める。また、政策一覧の中でスタンスを「やや賛成」や「やや反対」と明記されているものがあつた。本研究では、「やや」という文字を削除し、「反対」、「反対ではない」の2つにまとめた。そして、長期間での一貫性の有無を研究するために、2021 年度と 2022 年度のどちらの政策一覧にも掲載されていた5つの政策に絞って実験する。5つの政策は付録の図 5 に記載する。

一貫性のスコアの算出 3.3 節で、議員の発言を要約し、3.4 節で要約した文章と最も近い政策を得られるのだが、その政策のスタンスと 3.2 節で発言を分類したものが同じかどうかでスコアリングを行う。具体的に、今回は $score = \frac{\text{推論とマニフェストが同じだった数}}{\text{国会でマニフェストに関する話題が上がった数}}$ で計算を行った。この score を一貫性と呼び、選挙公約などで実現すると話していたことが、実際にどれだけ行動に移しているかを表した数値として見ることができる。本研究では、実験結果として政党ごとの政策に

2) <https://go2senkyo.com/>

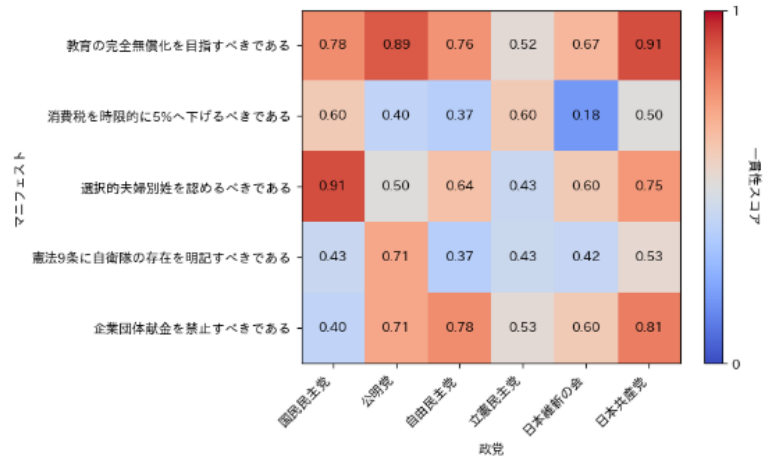


図3 ヒートマップによる一貫性の可視化

対する一貫性を分析することとする。

4 評価

表2に分類モデル, 3に要約モデルの評価を載せている。それぞれベンチマークとして, Fine-Tuningを行わなかった事前学習済みモデル (PLM) とモデルを使用せず全て反対でない (All-pros) とした場合を設定した。分類問題の評価に適合率, 再現率とF値を使用し, 要約問題の評価に ROUGE[16](表3ではR), BLEU[17]を使用した。

表2と表3に検証結果を記す。数値から分かるように, 上手く学習ができていたと考えられる。分類モデルに関して, 「全て反対ではない」と回答すると再現率が1.0になるのに対し, Fine-Tuningしたモデルは0.96であった。一見劣っているように思えるが, 適合率をも比較すると, 結果的にFine-Tuningしたモデルの方がスコアが良いように思える。

また, 一貫性があるかどうかを可視化したものが図3である。この図から分かることとして, 教育の無償化に公明党と日本共産党は一貫性が強くあると見られ, また夫婦別姓問題に関しては国民民主党に一貫性が強くあると見られる。すなわち, 公明党と日本共産党は少なくとも2021年, 2022年の2年間では教育の無償化という政策に懸命に取り組んでいたと考えられる。その裏付けとして, 公明党のHP[18]によると, 2019年から教育の無償化という点に力を入れていることを確認できた。一方で, 日本維新の会が消費税の減税に関して, 一貫性が見られなかったという点がある。この点を言い換えると, 「日本維新の会は消費税の減税を本当に推しているのかは不明である」と結論づけることができる。日本維新の会の政

表2 分類モデルの評価

	Pre.	Rec.	F1
PLM	0.25	0.03	0.06
All-pros	0.54	1.0	0.70
Fine-Tuning	0.60	0.96	0.74

表3 要約モデルの評価

	R-1	R-2	R-L	BLEU
PLM	0.4322	0.1844	0.3924	0.0933
Fine-Tuning	0.5470	0.3885	0.6056	0.1205

策提言 [19]を確認すると, コロナ対応のための短期の景気対策として消費税減税を提言しており, 長期的な成長戦略として消費税を8%に減税することを提言している。

5 おわりに

本研究では, 政党ごとの政策に対するスタンスの一貫性という観点に着目して, 国会議事録を分析した。代表的な5つの政策に絞って実験した結果, 一貫して取り組んでいられるように思える政党が見られた一方で, そうでないと思われるような政党をも見つけることができた。

今後の展望として, データ量の拡張, 国会議事録のより良いデータセットの作成などがあげられる。今回は国会議事録の中でも一部を抽出した発言から一貫性という観点を検証したが, 議事録全体を入力できるように拡張すると結果が変わるかもしれない。そのためには, データのクリーニングが必要になってくると考える。また, 2020年より大流行した新型コロナウイルス感染症であるが, この前後で感染症に対するスタンスなども変わっているかと思う。そのような転換期を考慮して変遷を分析することも必要になってくると考える。

参考文献

- [1] 総務省. 国政選挙の年代別投票率の推移について, (2023-12 閲覧). https://www.soumu.go.jp/senkyo/senkyo_s/news/sonota/nendaibetu/.
- [2] NHK. なぜ若者たちは投票に行かないの? 投票率を上げるには?, (2023-12 閲覧). <https://www.nhk.or.jp/hokkaido/articles/slug-n12893146ddae>.
- [3] 猫本隆哲, 秋葉友良, 増山繁. 議題への明示的な賛否表現を含まない議員発言を対象とした議会議事録のスタンス分類. 言語処理学会第27回年次大会, pp. 669–673, 2021.
- [4] 大杉了斗, 秋葉友良, 増山繁. 議事録の質疑応答の対応関係に基づくクラスタリングによる議題の抽出. 言語処理学会第27回年次大会, pp. 1492–1496, 2021.
- [5] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks, 2014.
- [6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. **Advances in neural information processing systems**, Vol. 30, , 2017.
- [7] Adrien Bougouin, Florian Boudin, and Béatrice Daille. TopicRank: Graph-based topic ranking for keyphrase extraction. In Ruslan Mitkov and Jong C. Park, editors, **Proceedings of the Sixth International Joint Conference on Natural Language Processing**, pp. 543–551, Nagoya, Japan, October 2013. Asian Federation of Natural Language Processing.
- [8] Rui Meng, Sanqiang Zhao, Shuguang Han, Daqing He, Peter Brusilovsky, and Yu Chi. Deep keyphrase generation. In Regina Barzilay and Min-Yen Kan, editors, **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 582–592, Vancouver, Canada, July 2017. Association for Computational Linguistics.
- [9] Hugging Face. cl-tohoku/bert-base-japanese-v3, (2023-12 閲覧). <https://huggingface.co/cl-tohoku/bert-base-japanese-v3>.
- [10] NTCIR-14. QA Lab PoliInfo, (2023-12 閲覧). <https://poliinfo.github.io/>.
- [11] Hugging Face. sonoisa/t5-base-japanese-title-generation, (2023-12 閲覧). <https://huggingface.co/sonoisa/t5-base-japanese-title-generation>.
- [12] Fundamental AI Research. faiss, 2017. <https://github.com/facebookresearch/faiss>.
- [13] 選挙ドットコム. 衆院選 2021 の各党公約を一覧で比較!, (2023-12 閲覧). <https://go2senkyo.com/articles/2021/10/25/63878.html>.
- [14] 選挙ドットコム. 参院選 2022 の各党公約を一覧で比較!, (2023-12 閲覧). <https://go2senkyo.com/articles/2022/07/07/69339.html>.
- [15] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. **IEEE Transactions on Big Data**, Vol. 7, No. 3, pp. 535–547, 2019.
- [16] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In **Text Summarization Branches Out**, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [17] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In Pierre Isabelle, Eugene Charniak, and Dekang Lin, editors, **Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics**, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.
- [18] 公明党. いよいよ始まる「3つの教育無償化」, (2023-12 閲覧). <https://www.komei.or.jp/content/p34280/>.
- [19] 日本維新の会. 政策提言 維新八策, (2023-12 閲覧). <https://o-ishin.jp/about/seisakuteigen/>.

A 付録

表4 よく見られた定型文の一部

ご質問ありがとうございます
分かりました
御答弁ありがとうございます
以上です
終わります
ご指摘のとおりでございます
動議を提出いたします
この際、一言ご挨拶申し上げます

表5 本研究で指標とした5つの政策

消費税を時限的に5%へ引き下げるべきである
憲法9条に自衛隊の存在を明記すべきである
教育の完全無償化を目指すべきである
選択的夫婦別姓を認めるべきである
企業団体献金を禁止すべきである

表6 国会議事録の発言数

	抽出前		抽出後	
	2021年	2022年	2021年	2022年
自由民主党	24557	26359	8214	10034
国民民主党	4825	5271	1563	1345
立憲民主党	20002	21589	6509	6002
公明党	6361	5768	1459	2393
日本維新の会	8243	5256	3134	2218
日本共産党	7059	7018	2301	2091