

## 国会答弁の言語特徴量と印象の相関調査 Correlations between impression and language feature quantity from text mining of Diet speech.

御崎 淳十  
Atsushi Misaki

成見 哲十  
Tetsu Narumi

### 1. はじめに

国政選挙や地方選挙は、候補者や政党への投票を通じて有権者が主権を行使し、間接民主制を実現する重大な行事である。選挙で選ばれた議員は議会への法案の提出、審議、採決や、予算の議決、執行機関の監視などを行い、それは国のあり方や私たちの生き方を左右する。このため、有権者の主権の行使には重大な責任が伴い、有権者はよく考えて投票先を決定すべきである。

しかし、例えば 2012 年の衆議院小選挙区の投票率が 59.32% と戦後最低を記録するなど、投票率は徐々に下がっている。投票を棄権した理由として、「仕事や重要な用事があった」「適当な候補者がいなかった」「候補者の人物像がよくわからなかった」などの理由が上位に来ている [1]。

インターネットの発達した現代では、知ろうと思えば候補者の情報を集めることは容易である。個別の候補者の情報は勿論、国会答弁の動画や議事録などの一次情報には誰でもアクセスすることが出来る。しかし、実際に有権者が触れる情報は、看板・選挙公報・新聞・テレビ・ラジオ・街頭演説などが多く、必ずしも有権者の要望をかなえてくれる訳ではない。例えば、看板や選挙公報はスペースが限られていて、有権者は候補者の名前、顔、公約の概要しかわからない。報道は直接情報源に当たれないので、その情報をどこまで信用していいのかわからない。街頭演説や講演会で直接候補者に会って話を聞くのは、候補者の政策や人柄を知る上で良い方法であるが、有権者の都合や手間の問題がある。国会答弁や議事録は膨大で、調査に時間がかかる。

そこで、選挙で有権者が候補者について信頼できる十分な量の情報を得るために、議会の議事録を活用することを提案する。議事録は一次情報なので信頼できるが、情報が膨大過ぎて候補者の情報だけを的確に取り出すことが難しかった。

膨大な情報から候補者に対する信頼感などの印象について短時間で知ることが出来れば、従来政策だけに偏りがちだった候補者に関する情報の質を上げることができると考える。

本研究では、候補者の国会答弁から、有権者が選挙に必要な情報だけをテキストマイニングを用いて抽出することを目指す。

### 2. 既存研究

上神貴佳らは、日本での有権者の各政策への考え方と

†電気通信大学大学院 情報理工学研究科 情報・通信工学専攻

政党のマニフェストをマッチングするシステムを開発した [2]。ユーザがいくつかの政策への賛否を回答すると、その考え方に近い政党名が出力される。

小磯花絵はテキストを読んで受ける印象を 3 つに分類し、それをテキストマイニングによって影響を与える言語の特徴を示した [3]。評定実験で因子分析を行い、テキストを「スタイル」「抑揚・リズム性」「整然性」に分類した。これらは、名詞、各種語種、常態・敬体、終助詞の出現率、文末の形態、文長、音転訛の出現率などが分類に影響を与えることを線形重回帰分析で示した。

本研究では、候補者本人の政策以外の情報を提供することで、有権者に幅広い選択肢を示す。また、国会議事録に特化したテキストマイニングを行い、選挙の役に立つ分類を目指す。

### 3. 提案システム

有権者が国会議事録から、候補者の印象的な答弁を知ることが目的とする。答弁の印象と議事録の言語特徴量の相関情報をもとに、候補者の答弁から有権者が必要な答弁を抽出する。

#### (1) 分類指標の策定

候補者の国会答弁を評価する指標を策定する。今のところ、答弁の正確性、共感性、安心感、好印象を 5 段階で評価することを考えている (表 1)。

#### (2) ラベル付け

被験者に国会の動画を見てもらい、分類指標をもとに答弁を評価する。国会の動画は衆議院、参議院が提供しているものを用いる。

#### (3) 言語特徴量の抽出

議事録からテキストマイニングのための言語特徴量を策定する。例えば、固有表現率、名詞率、動詞率、形容詞率、文長、文長変動率というように決める。そして、議事録から言語特徴量を抽出する。議事録は国会図書館が提供しているものを用いる。現在、演者の演説と演者に対する答弁の演説の、固有表現率、文末の形態変動率、「…という」といった説明をする言葉の出現頻度に注目している。

#### (4) 相関調査

被験者が評価したラベルと言語特徴量の相関を調べる。

表 1. 答弁の分類指標。

指標	段階
正確性	5 段階
共感性	
安心感	
好印象	

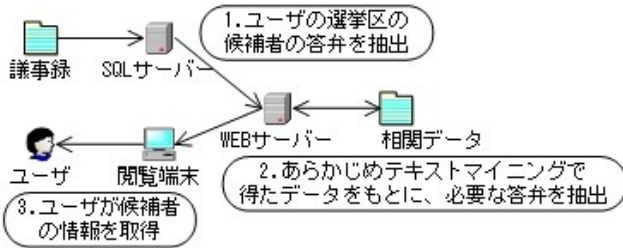


図 1. ユーザへ議事録を提供する流れ.

候補者の答弁から、必要な答弁のみを抽出する。

表 2. 議事録の SQL の設計.

(a) 演説テーブル

URL	SpeechNumber	Speaker	Speech
text	int	text	text

(b) 会議テーブル

URL	Kind	Title	Date	Start	End
text	text	text	date	time	time

例えば、重回帰分析などを用いて評価する。被験者全員のデータで相関が得られない場合、被験者をクラスタリングしてから再度相関を検証する。

(5) 有権者への提供

相関調査で得られた値をもとに、議事録から評価の高い答弁を抽出し、有権者に提供する(図 1)。

#### 4. 予備実験

(1) 議事録の取得

国会図書館の WEB ページから議事録を取得し、SQL サーバーに格納した。表 2 に SQL の設計を示す。第 145 回から第 183 回の衆議院、参議院、両院協議会の議事録(誤字、脱字のあった 4 会議を除く)を格納した。また、演者の名前は肩書きを外しフルネームに置換した。

(2) ラベル付け WEB ページの作成

被験者がラベルを入力する WEB ページを作成した(図 2)。内部で SQL サーバーに接続していて、議事録の取得とラベルの送信を行っている。ラベル付けのときは、国会動画とこの WEB ページを使用する。

(3) 被験者 1 人でのラベル付けと重回帰分析

現在、ラベル付け作業を行っている。

#### 5. 今後の作業と評価方法

言語特徴量抽出やラベル付けのためのシステムはほぼ完成しているが、被験者による調査や検証が出来ていない。答弁を分類する指標の妥当性の調査や、動画と議事録の言語特徴量との相関を調べるためのラベル付け作業が必要である。それらが取得できシステムが完成した後には、実際の投票前に被験者による評価実験を計画している。

(1) 分類指標の策定

選挙で候補者のどこに注目すればよいかを調査して決める。国会の動画の視聴、議員や有権者への質問を実施する。分類指標に重複がないか、被験者 10 人に動画で答弁を 1 時間分見てラベル付けをしてもらい、検証する。

(2) ラベル付け

syugin 第183回国会外務委員会第6号 2013-0

菊田真紀子

昨年、GSOMIAを締結しようとし、力を推進していくためのさまざまな努力、1 初め複数の閣僚が靖国神社を参拝し、安倍が整っていた外相の訪日を取りやめました。1 す。北朝鮮の挑発行為をやめさせるために中国は北朝鮮を援助する立場にあるからでことは、外交上得策だったと考えますか。1

正確性 5 4 3 2 1

共感性 5 4 3 2 1

安心感 5 4 3 2 1

印象の良さ 5 4 3 2 1

図 2. ラベル付け WEB ページの画面設計.

上部に会議情報、演者、演説を表示し、下部にラベル付け、送信、次の演説に移るボタンを配置した。

表 3. 評価項目.

評価項目	回答方法
候補者の政策	正誤
候補者への親近感	5段階
国会で話し合われていること	正誤
議会への親近感	5段階

被験者 10 人に、分類指標をもとにしたラベル付けをしてもらう。動画で答弁を 2 時間分見てもらう。

(3) 言語特徴量の抽出

フリーのソフトウェアを改良するなど既存の手法で答弁の言語特徴量の抽出を行う。

(4) 相関調査

ラベルと言語特徴量の相関を、重回帰分析など既存の方法を用いて検証する。

(5) 評価

候補者について知った情報の数、国会で何を話し合われているかの項目についての評価を、政見放送、新聞記事、国会中継、国会議事録と比較する(表 3)。10 人の被験者で、システムを評価する。選挙の投票を目的に、このシステムを 30 分使ってもらう。別の被験者に、ほかの方法を使ってもらい比較する。

#### 参考文献

- [1] 総務省選挙部, 目で見る投票率(平成 24 年 3 月), [http://www.soumu.go.jp/main\\_content/000153570.pdf](http://www.soumu.go.jp/main_content/000153570.pdf), 2012.
- [2] 上神貴佳, 堤英敬, 投票支援のためのインターネット・ツール-日本版ポートマッチの作成プロセスについて-, 選挙学会紀要, 2008.
- [3] 小磯花絵, テキストの多様性をとらえる分類指標-体系化の試み-, コーパスとテキストマイニング Corpus and Text Mining, 共立出版株式会社, 2012.