

国会会議録に基づく短命大臣の特徴分析

大南勝

s1620760@u.tsukuba.ac.jp

掛谷英紀

kake@iit.tsukuba.ac.jp

筑波大学

概要 近年、国務大臣や知事など、政治的要職につく人物が、不用意な発言や不祥事の発覚等により、任期途中にその職を追われるケースがしばしば見られる。安定した政治を行うためには、より慎重な人選を行うことが必要とされている。実際、首相が閣僚人事を行う際には、入閣候補者について「身体検査」と呼ばれる身辺調査が行われている。しかし、その詳しい手法については明らかにされていない部分が多い。そこで、本研究では、Web 上に公開されている国会会議録を言語資源に用い、長期的に大臣職を務めあげた人物と、大臣就任後舌禍や不祥事によって辞任した人物の特徴を機械学習し、その分類器の作成を試みる。また、その分類結果の正当性を機械学習のクロスバリデーションの正答率で評価するとともに、短命大臣に特徴的な素性について抽出・分析を行う。

1. はじめに

近年、政治の世界において、重要ポストに就いた人物が問題発言やスキャンダルによって任期途中で辞任することが相次いでいる。東京都では、2人の知事が立て続けに任期途中で辞任するという異例の事態が生じている。また、国務大臣についても、不用意な発言や不祥事の発覚等により、任期途中で辞任に追い込まれることもしばしば見られる。

その一方、内閣改造でも重要ポストに再任され、その職務を長期にわたって務めあげる人物もいる。こうした人選を全てのポストに対して実施できれば、政治はより安定したものになる。

閣僚人事は、その後の内閣支持率を大きく左右するものであるため、組閣時にいわゆる「身体検査」が行われることは広く知られている。具体的には、閣僚候補者の過去の政治活動と発言を調査し、直近の2回の選挙における収支報告書を点検するなどの作業が行われている。しかし、その詳しい手法については明らかにされていない部分が多い。

そこで、本研究では Web 上で公開されている国会会議録をもとに、機械学習を用いて、長期的に大臣職を務めあげた人物と、大臣就任後舌禍や不祥事によって辞任した人物の分類を試みる。

2. 研究手法

これまでも国会会議録を対象として自然言語処理

を応用した研究はいくつかあるが[1-4]、本研究では、国会会議録に収められている国会答弁を用い、国務大臣を長く続けた人物(長期大臣)と、任期途中で辞任した人物(短期大臣)を判別するシステムの構築を試みる。

まず、国会会議録検索システム[5]から入手した国会会議録の文書データを、形態素解析ツール MeCab[6]を用いて形態素ごとに分割する。本研究では、機械学習に用いる形態素を名詞・動詞・形容詞に限る。また、動詞・形容詞についてはそれぞれ文章中で現れた表出形から活用情報を省いた基本形を素性として使用する。

このようにして抽出した素性をもとに、学習データおよびテストデータを作成する。その後、学習データに対して、各カテゴリ間の特徴を機械学習し、テストデータを使用して、開発したシステムの精度を算出する。精度を算出する際には、クロスバリデーション(交差検証)を用いる。以上の手順を図1に示す。

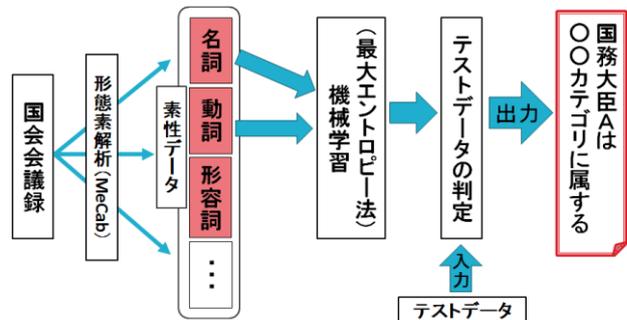


図1 システムの概要

3. 取得データ

3.1 国務大臣

本研究では、短期大臣を「内閣改造以外(舌禍・不祥事・引責等)で任期途中に大臣職を辞任した人物」と定義する。一方、長期大臣を「二度の内閣改造を乗り越えて同一職を務めた大臣」と定義する。分析対象は、昭和62年11月6日から平成28年8月3日の期間内に大臣職(内閣総理大臣は除く)に就いていた人物に限定する。

上記の定義にあてはまる人物を調査したところ、長期大臣は14名(延べ15名)、短期大臣は48名(延べ49名)が該当した。長期大臣について、氏名・役職・政党・在任日数をまとめたものを表1に示す。また、短期大臣については氏名のみを表2に示す。

表1 長期大臣

	氏名	役職	政党	在任日数
1	川端達夫	総務大臣	民主党	396
2	細野豪志	環境大臣	民主党	396
3	玄葉光一郎	新しい公共」担当大臣	民主党	452
4	野田佳彦	財務大臣	民主党	452
5	枝野幸男	経産大臣	民主党	472
6	玄葉光一郎(2)	外務大臣	民主党	482
7	藤村修	内閣官房長官	民主党	482
8	鹿野道彦	農水大臣	民主党	627
9	北澤俊美	防衛大臣	民主党	717
10	自見庄三郎	金融担当大臣	国民新党	728
11	中山太郎	外務大臣	自民党	818
12	宮澤喜一	大蔵大臣→財務大臣	自民党	1002
13	麻生太郎	財務大臣	自民党	1317
14	岸田文雄	外務大臣	自民党	1317
15	菅義偉	内閣官房長官	自民党	1317

表2 短期大臣

1	長谷川峻	18	柳田稔	35	中村正三郎
2	中山成彬	19	佐田玄一郎	36	松岡利勝
3	遠藤武彦	20	江藤隆美	37	福島瑞徳
4	鉢呂吉雄	21	松下忠洋	38	鳩山邦夫
5	永野茂門	22	額賀(2)	39	亀井静香
6	佐藤孝行	23	中西啓介	40	田中眞紀子
7	山下徳夫	24	中川秀直	41	松本龍
8	額賀福志郎	25	越智通雄	42	瓦力
9	田中慶秋	26	中川昭一	43	島村宜伸
10	太田誠一	27	前原誠司	44	三塚博
11	久世公堯	28	西川公也	45	橋本龍太郎
12	原田憲	29	久間章生	46	田名部匡省
13	桜井新	30	大島理森	47	宮澤喜一
14	小淵優子	31	船田元	48	甘利明
15	松島みどり	32	中島衛	49	福田康夫
16	赤城徳彦	33	奥野誠亮		
17	田沢智治	34	小泉純一郎		

3.2 国会議事録

本研究では、国会会議録検索システム検索用APIを利用して、分析対象の大臣の発言を収集する。なお、国会会議録の構造は図2に示す形式となっている。

○国務大臣(小泉純一郎君) 基本においては、障害者にとって住みやすい町をつくっていくと。そこで郵政省といいたしましても、今国会にはテレビの字幕放送、解説放送そして車いす用公衆電話ボックスの設置等、そういう助成等を図るための法案を提出して、去る五月十八日に成立いたしました。
各省協力しながら、いわゆるノーマライゼーションといいますが、完全参加と平等という、障害者も健常者と同様にいろんな事業に参加していく、障害者に優しい町づくりに郵政省としても協力していきたい、そう思っております。

図2 国会会議録の構造

先述の通り、長期及び短期大臣の過去の国会における発言から各カテゴリの特徴分析を試みる。3.1節で条件にあてはまった長期・短期の各大臣について、国会会議録に収められている発言の収集対象期間を次のように設定する。

長期大臣については、発言収集期間を大臣就任日から退任日までとする。一方、短期大臣の発言収集期間は大臣就任日から不祥事発覚日前日までとする。これは、不祥事発覚後の国会での追及に対する答弁に現れる特徴的な発言の影響を排除するためである。ただし、短期大臣に関して不祥事発覚日が不明な場合は退任日までとする。なお、衆参両院の国会会議録に含まれる全ての発言を対象とする。国会会議録データは図2のようになっているため、「○役職(氏名)」の部分を取り除き、それ以降をその人

物の発言として抽出する。

以上の基準で発言の収集を試みたところ、長期大臣は表 1 に載せた 15 名全員の発言が取得された。一方短期大臣については在任期間が数日から数十日の人物も含まれているため、表 2 に載せた 49 名のうち 11:久世公堯から 49: 福田康夫の計 39 名の発言のみ取得可能であった。取得発言件数の合計は長期大臣が 36490 件、短期大臣が 32569 件となった。これらのデータをもとに学習データを作成し、機械学習を行う。

4. 実験

4.1 データセット

3.2 節で収集した発言を、表 3 に示すグループを作成し、それに基づいて機械学習を行う。

表 3 データセット □内は発言数

	長期	短期
財務 1	麻生太郎 [4468]	中川昭一 [3904] 橋本龍太郎
財務 2	野田佳彦 [2236]	三塚博 [2249]
財務 3	宮澤喜一 [4811]	宮澤喜一(2) [4221]
外務 1	玄葉光一郎(2) [2122]	前原誠司 [821]
外務 2	中山太郎 [3157]	田中眞紀子 [2622]
官房長官 1	菅義偉 [2122]	福田康夫 [4231]
官房長官 2	藤村修 [639]	中川秀直 [193]
防衛	北澤俊美 [1822]	中西・瓦・額賀・久間 [2166]
経産系	枝野幸男 [1838]	小淵優子 [1758] 甘利明
農水	鹿野道彦 [2202]	赤城・西川・島村・松岡・大島 [2437]
総務	川端達夫 [1911]	江藤隆美 [1542] 鳩山邦夫
金融系	自見庄三郎 [771]	越智・久世・松下 [532]
その他 1	玄葉光一郎 [362]	奥野誠亮 [524]
その他 2	細野豪志 [1682]	桜井・佐田・福島・松本 [1801]

機械学習を行う際には、各カテゴリの発言件数の差が大きいと、発言件数の多いカテゴリに判定される確率が大きくなる可能性がある。そこで本研究では短

期大臣を複数人組み合わせるべく長期大臣の発言数に揃えてデータセットを作る。また大臣の役職による判定結果の偏りを少なくするため、同じ、もしくは近い役職の長期大臣と短期大臣を対にして学習を行う。また、クロスバリデーションはリーブ・ワン・アウト(ジャックナイフ法)で行う。

4.2 機械学習

上記のデータセットに基づき、最大エントロピー法による機械学習を行った結果を表 4 に示す。

正解率は 75.0%となり、国会の発言をもとに長期大臣及び短期大臣を言い当てることある程度可能であることが分かる。

表 4 判定結果

	再現率	適合率	長期 判定数	短期 判定数
長期大臣	71.4%	76.9%	10	4
短期大臣	78.6%	73.3%	3	11

4.3 単語出現頻度

最大エントロピー法を用いた機械学習では、機械学習の結果算出される α 値をもとに、カテゴリを判定する際の手掛かりとなる素性を知ることができる。しかし、最大エントロピー法による学習では、各素性が文書データに出現するかしないかの 2 値情報しか与えられないため、素性の出現頻度情報は考慮されない。今回のように、大量の発言を人物ごとにまとめている場合には、各素性の出現の有無だけでなく、出現頻度情報も有益な情報となる。

そこで、長期大臣と短期大臣のそれぞれについて、各素性の出現頻度を調査し、ヒストグラムを作成する。ここでは取得したすべての発言における使用頻度を調査した。両カテゴリの合算で出現頻度が高い素性から、カテゴリによって出現頻度が大きく異なる素性を抽出する。長期大臣でより多く見られる素性を表 5、短期大臣でより多く見られる素性を表 6 に示す。

長期大臣で特徴的な素性である「御存じ」は「御存じのとおり」、「是非」は「是非ご理解いただきたい」といった使われ方が主にされている。また「控える」は「発言を控えさせていただきます」といった発言の中で使われている。一方、短期大臣で特徴的な素性である「尽くす」は「全力を尽くす」、「頑張る」「一生懸命」は「頑張っているところでございます」「一生懸命努力はさせていただいている」「一生懸命頑張ってまいりました」といった使われ方が主にされている。このことから長期大臣は相手を取り込みながらうまく納得

させるような発言が目立ち、不用意な発言は普段からしないように気をつけていることがうかがえる。一方、短期大臣は「頑張る」「一生懸命」のように自分の努力を一方的にアピールする発言が目立っており、自己中心的で聞き手への配慮に欠ける面があることがうかがえる。

表 5 長期大臣の特徴的な素性の一部抜粋

単語	長期大臣 出現頻度	短期大臣 出現頻度
認識	5682	2217
皆さん	3666	1805
確認	2625	1087
様々	2517	609
御存じ	2032	461
引き続く	1176	377
是非	1130	197
検証	1015	280
大切	684	298
控える	680	207

表 6 短期大臣の特徴的な素性の一部抜粋

単語	長期大臣 出現頻度	短期大臣 出現頻度
対処	999	1727
くれる	756	1461
くださる	610	1394
論議	93	1778
尽くす	507	1036
お互い	543	906
頑張る	314	941
なさる	263	847
願う	229	837
一生懸命	274	689

5. おわりに

本研究では国会会議録に収録されている、国務大臣の発言をもとに、長期的に大臣職を務める人物と任期途中で辞任することに至った人物の特徴分析を試みた。機械学習の結果、それぞれの人物群の分類がある程度可能であることが分かった。また、それぞれの人物群の発言について一定の特徴を見出すことに成功した。しかし、現段階ではデータセットが各カテゴリについて 14 個と少ない。今後は、大臣就任期間の発言のみならず、大臣就任以前の議員時代

の発言を収集する、長く首相を務めた人物と短期で首相を退いた人物をデータに含める、また国会議員出身の知事についても同様にデータを集める等の作業により、学習データ数を増やした分析を行う必要がある。また、頻度情報を反映するため、機械学習にサポートベクターマシンを導入するなどして、判定精度の向上を試みることも今後の課題である。

参考文献

- [1] 掛谷英紀, 尾崎正宗, 佐藤裕也 (2015): 国会における議員の発言の自動要約システム, 言語処理学会第 21 回年次大会
- [2] 尾崎正宗, 掛谷英紀 (2014): 国会会議録の主張文取り出しおよびその要約, 言語処理学会第 20 回年次大会発表論文集
- [3] 東宏一, 村田真樹, 掛谷英紀 (2010): 国会議事録を教師信号とする大学のイデオロギー分布推定, 言語処理学会第 16 回年次大会発表論文集
- [4] 畑中允宏, 村田真樹, 掛谷英紀 (2009): 新聞社説・国会議事録に基づく言論のイデオロギー別分類, 言語処理学会第 15 回年次大会発表論文集
- [5] 国会会議録検索システム <http://kokkai.ndl.go.jp/>
- [6] MeCab: Yet Another Part-of-Speech and Morphological Analyzer
<http://mecab.sourceforge.net/>