

国会会議録に基づく「チルドレン議員」の特徴分析

掛谷英紀

大南勝

kake@iit.tsukuba.ac.jp

s1620760@u.tsukuba.ac.jp

筑波大学

概要 本稿では、一方的なマスコミ報道の追い風に乗って大量当選したいわゆる「チルドレン議員」について、一期で終わってしまった人物と、その後議員を続けることができた人物の間で、一期目における国会での発言にどのような違いが見られるかを分析する。その結果、その後議員を続けられた人物は国会での質問回数が多いことが分かった。また、最大エントロピー法による学習の結果、一期で消えてしまった人物の発言には、漢字二字熟語で表される言葉の使用頻度が少ないこと、尊敬語・謙譲語などの丁寧な表現が少ないこと、損得勘定に関する表現の使用が多いことが分かった。

1 はじめに

2017年7月2日に投開票が行われた東京都議会議員選挙では、小池百合子知事が当時代表を務めていた都民ファーストの会から55人が当選したが、そのうち初当選組、いわゆる「小池チルドレン」と呼ばれる議員は39人にも及んだ。

マスコミの報道の追い風によって、政治経験のない人が「チルドレン議員」として登場することは、これまで何度も発生している。たとえば、2005年衆議院議員総選挙における「小泉チルドレン」、2009年衆議院議員総選挙における「小沢(鳩山)チルドレン」はその代表例である。しかしながら、これらのチルドレン議員の中には、当選後議員としての資質が疑われるような行動をする人も少なくない。実際、過去のチルドレン議員には、その後の選挙で全く当選のない人物が多くいる。その一方で、その後政治に長く携わる人物もいる。

筆者らは、これまで日本の国会会議録を言語資源とした研究に取り組んでいる[1-4]。国会議員の資質に関わる分析として、大南らは問題発言や不祥事によって任期中で辞任した短命大臣の特徴を調べる研究を行っている[5]。これまで、短命大臣の国会発言について、ある程度の特徴は得られているが、大臣のみを対象にすると、人数が少ないため、国会議員のより一般的な資質を見るには不十分な点も多かった。

そこで、本研究では、マスコミ報道の追い風に乗って大量当選したチルドレン議員について、一期で終わってしまった人物と、その後議員を続けることができた人物の違いを国会会議録をもとに分析することで、国会議員としての資質に関するより一般的な特徴を得ることを目的とする。

2 手法

本研究では、国会会議録に収められている国会質問を言語資源として、教師あり学習によって文書を分類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[6]から国会会議録検索システム検索用API[7]を用いて国会会議録の文書データを入手する。

次に、取得した文書データを形態素解析ツールChaSen[8]を用いて形態素ごとに分割し、素性データを決定する。このようにして抽出した特徴語から訓練データおよびテストデータを作成する。その後、訓練データを用いて、各カテゴリ間の特徴を機械学習し、その学習したモデルにテストデータを入力して判定結果を得る。

機械学習には最大エントロピー法を用いた学習と判別分析法を用いた学習をそれぞれ実装し、分類性能の比較を行う。精度を算出する際には、共にクロスバリデーション(交差検証)を行う。本研究では最大エントロピー法を用いた機械学習のプログラムとしてmaxent[9]を使用する。

3 小沢(鳩山)チルドレンの分類

本節では2009年8月30日に執行された第45回衆議院議員総選挙において、民主党から出馬し初当選を果たした人物、いわゆる小沢(鳩山)チルドレンを対象とした分類を行う。この選挙では鳩山由紀夫率いる民主党が圧勝し、自民党から政権を奪取したが、その際に多くのチルドレン議員が生まれている。この選挙以前に参議院議員を経験していた人物は除き、民主党から衆議院議員に初当選した人物は計139名であった。そのうち、その後の選挙で当選のない「消えたチルドレン」は計109名に及ぶ。首長や地方議員に転身した人物は5名、またその後の

選挙で参議院を含めて国政に復帰した人物は計 25 名で「生き残りチルドレン」は計 30 名である。なお、これらの数字はいずれも、2017 年 7 月時点での集計結果である。

最初に、各人物が当選日以降、次の解散までに質問に立った国会の会議数を調査する。「消えたチルドレン」計 109 名の発言国会会議数の平均は 10.0、標準偏差は 6.7 となった。一方、「生き残りチルドレン」計 30 名の平均は 13.0、標準偏差は 5.0 となった。ここで、両カテゴリ間での発言会議数に有意差があるか検証するため t 検定を行った。その結果、 $p = 0.025$ となり、有意水準 5% で、「消えたチルドレン」は「生き残りチルドレン」と比べて国会質問に立つ回数が有意に少ないことが示された。このことから、国会質問への積極的な姿勢が見られないチルドレン議員は、一度の当選のみに終わってしまう傾向が強いと考えられる。

次に、発言の収集および機械学習用のデータセットの作成を行う。本実験では、機械学習に用いる 1 人分の発言数を揃えるため、各議員について、一期目の国会発言のうち新しいものから最大 100 件の発言を収集する。また、学習に使用する各カテゴリの人数を揃えるため、以下のグループ分けを行う。

まず、「生き残りチルドレン」30 名のうち、発言会議数が 6 未満と極端に少ない人物 1 名（石井章）を分類対象から除き、計 29 名から成るグループ A を作成する。「消えたチルドレン」については、発言会議数が 9~12 と少ない人物を集めたグループ B、発言会議数が 11~16 と比較的多い人物を集めたグループ C を作成する。いずれのグループも議員 29 名を含む。なお、「消えたチルドレン」のうち発言会議数が 11~12 の計 16 名についてはグループ B、C のどちらにも含まれている。

このデータセットに対して、最大エントロピー法を用いた機械学習による分類を試みる。名詞・動詞・形容詞を素性として使用し、クロスバリデーションの分割数は 10 とする。グループ A の議員とグループ B の議員の分類結果を表 1 に、グループ A とグループ C の分類結果を表 2 に示す。

正解率の高いグループ A とグループ B の判定における、各カテゴリ 20 個の特徴的な (α 値の高い) 素性を表 3 に示す。「消えたチルドレン」(B) は「生き残りチルドレン」(A) に比べて漢字の二字熟語の使用頻度が少ない傾向があることが分かる。

次に、動詞のみを素性として使用した場合の分類を行う。使用素性を動詞に限定した場合の、グループ A とグループ B の分類結果を表 4 に示す。また、グループ A とグループ C の分類結果を表 5 に示す。

表 1 グループ A と B の分類結果
(名詞・動詞・形容詞を素性に使用)

	再現率	適合率	総数	判定 A	判定 B
A	65.5%	65.5%	29	19	10
B	65.5%	65.5%	29	10	19
総数	65.5%	65.5%	58	29	29

表 2 グループ A と C の分類結果
(名詞・動詞・形容詞を素性に使用)

	再現率	適合率	総数	判定 A	判定 C
A	48.3%	51.9%	29	14	15
C	55.2%	51.6%	29	13	16
総数	51.7%	51.7%	58	27	31

表 3 グループ A と B に特徴的な素性

A	B
含む, あり方, 担当, 設ける, どなた, 民間, 契約, どちら, 原因, 迎える, 総務, 制約, 行政, 挙げる, 仕事, 協定, 作業, 交渉, 関連, 集中	お尋ね, 教える, 生まれる, 申しわけ, 協議, 人口, 都市, 場所, 小宮山, 当初, 人員, 応援, 述べる, 危機, 研究所, 備える, 回る, うち, 違い, ども

表 4 グループ A と B の分類結果 (動詞のみ)

	再現率	適合率	総数	判定 A	判定 B
A	65.5%	67.9%	29	19	10
B	69.0%	66.7%	29	9	20
総数	67.2%	67.2%	58	28	30

表 5 グループ A と C の分類結果 (動詞のみ)

	再現率	適合率	総数	判定 A	判定 C
A	58.6%	63.0%	29	17	12
C	65.5%	61.3%	29	10	19
総数	62.1%	62.1%	58	27	31

機械学習に使用する素性を動詞のみにした場合にも、グループ A と B の分類が、グループ A と C の分類よりも高い正解率を示している。また、名詞・動詞・形容詞を使用素性とした場合と比べて、どちらの分類も正解率が向上している。これは、動詞のみに限定することで、所属委員会の違いによる使用する名詞の偏りが及ぼす影響が軽減されたためであると考えられる。各カテゴリの文章の違いがより顕著なグループ A とグループ B の判定における、各カテゴリ 20 個の特徴的な動詞を表 6 に示す。

表 6 グループ A と B に特徴的な動詞

A	B
含む, 育つ, 設ける, 聞こえる, 挙げる, 切れる, 取り組む, 応じる, 伺う, 任せる, しかる, 受けとめる, 変わる, 迫る, 減る, 失う, なさる, 過ぎる, 切る, 定める	生まれる, 述べる, 教える, 備える, 比べる, 動く, 抱える, たまる, 落ちる, 移る, 落ちる, 見つかる, 削る, くれる, おっしゃる, 救う, 重ねる, 尽くす, させる, やってくる

4 小泉チルドレンの分類

本節では自民党のチルドレン議員を扱う。ここでは 2005 年 9 月 11 日に執行された第 44 回衆議院議員総選挙で初当選を果たした自民党所属議員、いわゆる小泉チルドレンと呼ばれる人物を対象とする。この郵政総選挙とも呼ばれる総選挙では当時小泉純一郎が党首を務めていた自民党が 296 議席を獲得し、圧勝した。その際にも多くのチルドレン議員が生まれている。この選挙以前に参議院議員を経験していた人物は除き、自民党から衆議院議員に初当選した人物は計 83 名であった。そのうち、その後の選挙で当選のない「消えたチルドレン」は計 29 名である。一方、参議院議員や地方議員に転身した人物は 9 名、またその後の選挙で衆議院に復帰した人物は計 45 名である。なお、これらも 2017 年 7 月時点での集計結果である。

次に、発言した国会会議数を調査する。「消えたチルドレン」計 29 名の国会会議数平均は 12.8、標準偏差は 6.0 となった。一方、「生き残りチルドレン」計 54 名の平均は 20.7、標準偏差は 12.9 となった。ここでも、両カテゴリ間での発言会議数に有意差があるか検証するため t 検定を行ったところ、 $p = 0.0026$ となり、有意水準 1% で有意な差が見られた。小泉チルドレンにおいても「消えたチルドレン」は「生き残りチルドレン」と比べて国会質問に立つ回数がある程度少ないことが分かる。

続いて、発言の収集および、データセットの作成を行う。発言については前節同様、各人物、新しいものから最大 100 件の発言を収集する。本節では 2 つのデータセットを用意する。1 つ目は、のちに衆議院に復帰した「生き残りチルドレン」のうち発言会議数が 10~23 の人物を集めたグループとする。これをグループ D とする。グループ D に該当した議員は 26 名であった。2 つ目は、「消えたチルドレン」のうち発言会議数が 5 以上の人物 26 名を集め

たもので、これをグループ E とする。

最大エントロピー法を用いた機械学習により、グループ D と E の分類を試みる。名詞・動詞・形容詞を使用素性とした場合の分類結果を表 7 に、動詞のみを使用素性とした場合の分類結果を表 8 にそれぞれ示す。

表 7 グループ D と E の分類結果
(名詞・動詞・形容詞を素性に使用)

	再現率	適合率	総数	判定 D	判定 E
D	65.4%	53.1%	26	17	9
E	42.3%	55.0%	26	15	11
総数	53.9%	53.9%	52	32	20

表 8 グループ D と E の分類結果 (動詞のみ)

	再現率	適合率	総数	判定 D	判定 E
D	69.2%	64.3%	26	18	8
E	61.5%	66.7%	26	10	16
総数	65.4%	65.4%	52	28	24

本実験において、名詞・動詞・形容詞を使用した場合の正解率は 53.9%、動詞のみを使用した場合には 65.4% となり、11.5 ポイント向上した。ここでも、委員会の違いによる影響が軽減されたと考えられる。

名詞・動詞・形容詞を使用した判定における上位素性を表 9 に、動詞のみを使用した場合の上位素性を表 10 に 20 個ずつ示す。前節よりも顕著ではないものの、ここでも「消えたチルドレン」が漢字二字熟語の使用頻度が少ない傾向が見られる。また、「生き残りチルドレン」は「おっしゃる」「いらっしゃる」「伺える」など丁寧な言葉遣い多く見られることが分かる。さらに、「消えたチルドレン」は、「売る」、「もうかる」、「ふやす」など、商売の損得勘定に関する表現をよく使う傾向も見られる。

表 9 グループ D と E に特徴的な素性

D	E
代理, 着席, 退席, 質疑, 要望, 長, 来年, レベル, あれ, カバー, 共通, 形, 守る, 集中, 防衛, どれ, 出せる, 景気, 本来, 一生懸命	組織, お尋ね, 文化, 正確, 基づく, 売る, 経過, ぶり, バランス, 山, その他, これら, 条件, 向上, 報告, 生じる, 訴える, 際, 存じる, 見える

表 10 グループ D と E に特徴的な動詞

D	E
守る, 出せる, 成り立つ, 陥る, 通る, 思える, やってくる, 上げる, 教える, おっしゃる, 結びつく, こたえる, いらっしゃる, 起こす, 応じる, 載る, 広げる, 結ぶ, 結びつける, 伺える	売る, 基づく, 生じる, 存じる, もたらす, 当たる, 見える, 打ち出す, 戻る, 覚える, 沿う, 訴える, 承る, 許す, もうかる, 思い出す, 防ぐ, 頼る, ふやす, 死ぬ

5 チルドレン議員全体の分析

本節では、前節までで扱った小沢チルドレンと小泉チルドレンのデータを合わせて分類を試みる。グループ A とグループ D を合算した 55 名を「生き残りチルドレン」(グループ F)、グループ B とグループ E を合算した 55 名を「消えたチルドレン」(グループ G) として扱い、最大エントロピー法による機械学習を行う。名詞・動詞・形容詞を使用素性とした場合の分類結果を表 11 に、動詞のみを使用素性とした場合の分類結果を表 12 にそれぞれ示す。

小沢チルドレンと小泉チルドレンを合算した場合の分類においても、動詞のみを用いた場合の正解率が高く、64.6%となった。特徴的な動詞としては、「いらっしゃる」、「伺う」といった丁寧な表現が多い傾向がある。丁寧な表現について、 α 値をまとめたものを表 13 に示す。これらの α 値はグループ F から見た値である。また、比較のため丁寧語ではない「死ぬ」、「述べる」の α 値も括弧つきで載せている。このように、「生き残りチルドレン」は言葉遣いが丁寧な傾向があることが分かる。

表 11 グループ F と G の分類結果
(名詞・動詞・形容詞を素性に使用)

	再現率	適合率	総数	判定 F	判定 G
F	63.6%	59.3%	55	35	20
G	56.4%	60.8%	55	24	31
総数	60.0%	60.0%	110	59	51

表 12 グループ F と G の分類結果 (動詞のみ)

	再現率	適合率	総数	判定 F	判定 G
F	63.6%	64.8%	55	35	20
G	65.5%	64.3%	55	19	36
総数	64.6%	64.6%	110	54	56

表 13 グループ F と G を学習したときの丁寧表現の α 値 (グループ F から見た値)

素性	α 値	素性	α 値
いらっしゃる	0.585	おっしゃる	0.513
伺う	0.577	逝く	0.511
なさる	0.542	いただく	0.502
いただける	0.530	申し述べる	0.500
亡くなる	0.522	いたす	0.498
(死ぬ)	(0.475)	仰せつかる	0.496
申し上げる	0.519	召す	0.492
(述べる)	(0.390)	承る	0.468

6 おわりに

本研究では、チルドレン議員について、一期で終わってしまった人物と、その後議員を続けることができた人物の間の違いを分析した。その結果、その後議員を続けられた人物は国会での質問回数が多いことが分かった。また、最大エントロピー法による学習の結果、一期で消えてしまった人物の発言には、漢字二字熟語で表される言葉の使用頻度が少ないこと、尊敬語・謙譲語などの丁寧な表現が少ないこと、損得勘定に関する表現の使用が多いことが分かった。

ただし、分類の正解率は短命大臣の答弁を学習したときに比べてまだ低い。今後は用いる素性や収集するデータセットについて工夫をすることで、より精度の高い分析を行う必要がある。

参考文献

- [1] 畑中、村田、掛谷: 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, 2009
- [2] 尾崎: 国会会議録を用いた政党類似度マップの構築, 筑波大学理工学群工学システム学類卒業論文, 2013
- [3] 尾崎、掛谷: 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, 2014
- [4] 掛谷、尾崎、佐藤: 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, 2015
- [5] 大南、掛谷: 国会会議録に基づく短命大臣の特徴分析, 言語処理学会第 23 回年次大会, 2017
- [6] 国会会議録検索システム, <http://kokkai.ndl.go.jp/>
- [7] 国会会議録検索システム検索用 API, <http://kokkai.ndl.go.jp/api.html>
- [8] 奈良先端科学技術大学院大学 松本研究室 ChaSen <http://cl.aist-nara.ac.jp/>
- [9] Masao Uchiyama. Maximum Entropy Modeling Package. <http://www2.nict.go.jp/x/x161/members/mutiyama/software.html>, 2006