

Twitter データを用いた政治的論点の抽出と政治家分類

Politician classification based on political problems extracted from Twitter Data

野村 真由美

Mayumi Nomura

尾崎 知伸

Tomonobu Ozaki

日本大学 文理学部

College of Humanities and Sciences, Nihon University

In order to assist the decision making for political elections, in this paper, we propose a simple method to identify political problems from Tweets by politicians. Furthermore, using cluster analysis, we discover hidden communities among politicians by characterizing each politicians based on their attitudes on detected political problems.

1. はじめに

現在、日本の政治場面において、論点や立場が判然としない政党や団体の増加が散見される。また、政策や公約、政治活動のすべてが政治広報や選挙公報に記載されるわけではなく、各政治家の状態や思想が一覧できる資料を見つけることは非常に困難である。

これらのことを背景に、これまでに、ネット上の情報を用いて種々の論点を抽出し分析する技術が提案されている。テーマを問わないシステムとして、Web 上の意見や感想といった主観情報を用いて意思決定支援を行う OpinionReader[1] が挙げられる。OpinionReader では、賛否の両立するテーマに関する記事を入力し、テーマの何が問題となるのかという論点を抽出した上で、論点ごとの賛否を可視化する。また鷹栖ら [2] は、テーマに関する意見ツイートに対して係り受け解析を適用することで論点を抽出し、クラスタリングを通じて意見文进行分类する手法を提案している。同様に Jiang ら [3] は、ユーザの関連ツイートを考慮して意見文を肯定、否定、中立に分類する手法を、Kim ら [4] は文書を賛成反対の程度に応じて多値に分類する手法をそれぞれ提案している。

一方、論点抽出や意見分析の政治分野への応用も数多く試みられている。例えば、橋本ら [5] や東ら [6] は、Web サイトのデータを用いた議員分類を行っている。具体的には、Yahoo!JAPAN が運営する「みんなの政治」内に存在した「みんなの議員評価」と呼ばれるサービスを用い、一般ユーザが議員に対して行ったレビュー記事データから、各政党のイデオロギーの分析や議員の分類を行っている。また東ら [7] は、特定の政治テーマに対して明るい知識人のツイートを収集して賛成反対を分析し、その結果を基に著名議員のツイートから議員の立場を明らかにする手法を提案している。大南ら [8] は政治家ではなく一般のツイッターユーザに焦点を当て、政治に関係のない情報の混在する中から特定の政治テーマに関するツイートを集め、ユーザを分類している。また、暴力的なツイートを発するユーザ層のイデオロギーの相関についても検証を行っている。

これら既存研究の多くは、特定のテーマや論点を与えることを前提としている場合も少なくない。しかし、政治に詳しくない利用者にとっては、与えるべき論点を自主的に見つけることが自体が難しいといった問題が挙げられる。そこで本研究では、

政治家の日常の発信である Twitter^{*1} に焦点を当てる。議員のツイートから論点となりそうな単語を抽出するとともに、それらを整理し、また論点に対する立場から議員进行分类する。

2. データセット

ツイート集合収集時（2017 年 6 月 30 日）に国会議員を擁した政党 8 党に関し、各政党の公式アカウントがフォローしているアカウントからプロフィール等を参考に国会議員および関係者を推定し、分析の対象とした。分析対象のアカウント数は 413、総ツイート数は 753,841 件である。政党別の公式アカウントおよび分析対象アカウント数を表 1 に示す。

表 1: 分析対象政党とアカウント数

政党名	公式アカウント	対象アカウント数
自由民主党	jimin.koho	165
民進党	MinshintNews	101
公明党	komei.koho	47
日本共産党	jcp.cc	30
自由党	seikatsu1pr	26
日本のこころ	nipponkokoro	23
日本維新の会	osaka.ishin	21
社会民主党	SDPjapan	20

3. 政治的論点に関する賛否行列の構築

政治的論点を抽出するため、賛成語類語辞典と反対語類語辞典からそれぞれ賛成語 297 語、反対語 1,087 語を特定し、是非を問う単語群として利用する。賛成語・反対語の例を表 2 に示す。次に、各ツイートに対して係り受け解析ツール CaboCha^{*2} [9, 10] を適用して賛成語・反対語に係る名詞を特定し、これを政治的論点の候補とした。これらの候補から使用アカウント数が 5 未満の単語を排除し、最終的に 1,933 の政治的論点を特定した。特定された政治的論点の例を表 3 に示す。

各議員による各論点に対する態度を表現する、行数 413（＝対象議員数）、列数 1,933（＝政治的論点数）の賛否行列 M

連絡先: 尾崎 知伸, 日本大学 文理学部 情報科学科, 〒156-8550
東京都世田谷区桜上水 3-25-40, tozaki@chs.nihon-u.ac.jp

*1 <https://twitter.com/>

*2 <https://taku910.github.io/cabocha/>

表 2: 賛成語・反対語の例

賛成語		反対語	
賛成	受け入れ	反対	拒否
満足の意	好都合	不可	禁止
イエス	肯定	ノー	否定
異議なし	同意	異議あり	受け入れがたい
後押しする	承認	水を差す	抗議

表 3: 抽出された政治的論点の例

新たな活動	安倍政権	法案	賛成
最後	議員	時	一
様々	向け	多く	出席
声	考え	必要	対応
思い	今後	皆様	選挙
全く	党	議論	何
政治	何	応援	問題
方	参加	意見	政策
国	国会	中	質問
岡田代表	国民	今	私
見極め	選出	疑惑	続投
2 日	明記	国会招致	展望
ケア	歴史認識	乱暴	事実上
石川	母校	政権運営	おり
留学生	抱え	期間	いじめ
sns	第一	撤去	いくつか
再稼働	先行	細野豪志	現時点
国会内	防衛	修正案	二重国籍
内閣委員会	18 時	岩手	全国的
国会質問	不足	静岡	出産
4 項目	決着	動向	南シナ海
命令	イスラエル	お答え	配備
氏	保育所	天候	サポーター
自己	ガイドライン	半ば	見つけ
遠く	明後日	請願	神奈川
了承	四	北海道 5 区	入国
進展	金融	一時	未来のために
動い	内閣支持率	19 日	私学
大勢	おかげ	懇親	インターネット
後半	リアル	中途半端	供給
	熊本県	教育委員会	発議

表 4: 政治的論点・発言クラスタにおけるクラスタ中心

クラスタ数 8		クラスタ数 4	
賛否表現	発言表現	賛否表現	発言表現
態勢	こと	態勢	こと
ため	一	ため	一
特定	会	抗議	会
5 兆円	日本共産党	安保	反対
義務付け	疑惑		
安倍政権	辺野古		
中国	瀬谷		
辺野古	行使		

4. 政治的論点の分類

賛否行列表 M 及び発言行列表 U の各列に対し、コサイン類似度を用いた K-medoids 法を適用し、政治的論点及び利用される名詞をそれぞれグループ化する。得られたクラスタの中心 (Medoid) を表 4 にまとめる。

賛否行列表に対するクラスタ中心の一つである〈5 兆円〉は、防衛省の国家予算と年金運用にあたる損失額の両方を指す語である。また、〈安保〉は安全保障関連法案の略、〈特定〉は秘密法案にかかる語である。その他、〈安倍政権〉〈中国〉〈辺野古〉など、全体として論点として妥当な単語がクラスタ中心として選択されていることが分かる。

5. 政治家の分類

賛否行列表 M 及び発言行列表 U の各行に対してクラスタリング技術を適用し、対象となる議員をグループ化する。なお、クラスタリングにはコサイン類似度を用い、政党数である $K=8$ -medoids 法を用いた。得られたクラスタ中心 (議員) の所属政党と政党別のクラスタ所属率を表 5 に示す。

賛否表現に対するクラスタリング結果は、発言表現による結果と比べ、議員の所属政党のクラスタに関してばらつきが存在している。なお、議員の本来の所属を正解とした場合との RAND 尺度は、賛否表現で 0.71、発言表現で 0.76 であり、必ずしも政党内で各論点に対する賛否が分かれているわけではないこと、個々人の発言に沿った語句によりクラスタが形成されていることが伺える。

次に、賛否表現による政治家分類の結果を考察する。日本共産党、社会民主党所属議員の多くはクラスタ 3 に、公明党所属議員の多くはクラスタ 4 に、日本維新の会所属議員はクラスタ 5 に、それぞれ集まっていることが確認された。その一方で、政党内で最も多く分類されたクラスタと、そのクラスタの medoid の議員の政党が一致しているのはクラスタ 4 の公明党のみであり、政党という意味では、必ずしもクラスタ中心 (の議員) がそのグループを代表しているわけではないという結果となった。

またクラスタリング結果から、自由民主党と民進党の議員が各クラスタに分散していることが確認できる。これは、両党ともに議員数が多いため、党の内部において同一の論点の発信を行えないことを考えると妥当であると言える。クラスタ 7 に関しては、与党である自由民主党と日本のこころの議員が多く集まっているが、野党議員は非常に少ないことが分かる。この傾向は、同じく与党である公明党の議員が集まるクラスタ 4 においても確認できる。これらの結果から、与党と野党の論点

を考える。賛否行列表 M の第 i 行、第 j 列の要素 $M_{i,j}$ は、議員 i による政治的論点 j に対する態度であり、議員 i が論点 j に賛成した回数を $A_{i,j}$ 、反対した回数を $D_{i,j}$ としたとき、 $M_{i,j} = A_{i,j} - D_{i,j}$ と計算する。

一方比較のため、収集したツイートに含まれる名詞に着目した発言行列表 U を考える。発言行列表 U は、行数 413 (=対象議員数)、列数 1,621 (=名詞数) の行列であり、議員 i が単語 j を発信した回数を要素 $U_{i,j}$ として持つ。なお、名詞の抽出には解体素解析器 MeCab^{*3} [11] を用い、またすべてのアカウントで 100 回未満しか発信されなかった名詞は排除している。

賛否行列表 M と発言行列表 U に対し、共通して現れる名詞 (列) 数は 408 であった。また M にのみ現れる名詞数は 1,525、 U にのみ現れる名詞数は 1,213 であった。これより、両者の Jaccard 類似度は 0.130 となり、両者で捉えている特徴が大きく異なることが分かる。

^{*3} <http://taku910.github.io/mecab/>

表 5: 各政党別のクラスタ所属率

ID	クラスタ中心	自由民主党	民進党	公明党	日本共産党	自由党	日本のこころ	日本維新の会	社会民主党
賛否表現									
1	自由民主党	0.07	0.03	0.07	0.00	0.22	0.06	0.00	0.00
2	民進党	0.14	0.15	0.14	0.04	0.17	0.00	0.16	0.08
3	自由民主党	0.07	0.13	0.00	0.63	0.22	0.19	0.00	0.54
4	公明党	0.14	0.09	0.41	0.04	0.00	0.06	0.16	0.00
5	民進党	0.18	0.19	0.18	0.11	0.11	0.13	0.32	0.15
6	日本共産党	0.12	0.06	0.07	0.04	0.17	0.13	0.26	0.15
7	自由党	0.20	0.20	0.11	0.07	0.06	0.31	0.05	0.08
8	自由民主党	0.09	0.13	0.02	0.07	0.06	0.13	0.05	0.00
発言表現									
1	民進党	0.15	0.37	0.00	0.00	0.35	0.25	0.05	0.23
2	日本共産党	0.01	0.03	0.00	1.00	0.20	0.00	0.00	0.23
3	民進党	0.15	0.15	0.09	0.00	0.05	0.00	0.00	0.00
4	自由民主党	0.38	0.08	0.05	0.00	0.15	0.00	0.00	0.00
5	日本維新の会	0.00	0.00	0.00	0.00	0.00	0.25	0.84	0.00
6	自由民主党	0.01	0.24	0.00	0.00	0.05	0.25	0.00	0.38
7	公明党	0.10	0.02	0.82	0.00	0.05	0.13	0.00	0.15
8	自由民主党	0.20	0.10	0.05	0.00	0.15	0.13	0.11	0.00

発信の方向性が大きく異なることが考えられる。

各政党の少数派によって形成されているクラスタも確認できる。これは、論点の数が政党数に満たない、または複数の政党で論点を共有している結果とも考えられる。また、論点クラスタの中心（表 3）に、政治的ではない一般名詞が含まれており、政治家分類においても、（本当の意味での）論点だけでなく、（係り受け解析により抽出された）一般名詞が大きく関係していると考えられる。

6. おわりに

本研究では、議員および政治家関係者の発信するツイートを対象に論点を抽出し、得られた論点を用いて政治家の分類を行った。その結果、与野党間の差をある程度反映しながらも、政党には寄らない分類結果が得られることが示された。

今後の課題としては、政治的論点抽出の精度向上を目的に、政治分野特有の固有名詞を対象とした固有表現抽出技術の開発や、一般名詞の排除などが挙げられる。また、二重否定や反語表現など、単純な係り受け解析だけでは特定することが困難な論点への態度の決定についても、更なる考察を行う予定である。

参考文献

- [1] 佐々木 千晴, 藤井 敦, 石川 鉄也: 意思決定支援のための主観情報マイニング, 言語処理学会第 12 回年次大会発表論文集, pp.77–80 (2006)
- [2] 鷹栖 弘明, 内海 彰: 文脈を考慮した観点に基づく意見ツイートクラスタリング, 言語処理学会第 21 回年次大会発表論文集, pp.433–436 (2015)
- [3] L. Jiang, M. Yu, Mi. Zhou, X. Liu, and T. Zhao: Target-dependent Twitter Sentiment Classification, *Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp.151–160 (2011)
- [4] S.-M. Kim and E. Hovy: Determining the sentiment of opinions, *Proc. of the 20th international conference on Computational Linguistics*, Article No. 1367 (2004)
- [5] 橋本 悠, 掛谷 英紀: Web 上のレビュー記事のイデオロギー分析とその応用, 言語処理学会第 16 回年次大会発表論文集, pp.740–743 (2010)
- [6] 東 宏一, 橋本 悠, 掛谷 英紀: WEB 上の言語資源に基づく国会議員の分類, 言語処理学会第 17 回年次大会発表論文集, pp.476–479 (2011)
- [7] 東 宏一, 掛谷 英紀: 国会議員のツイッター分類とその応用, 言語処理学会第 18 回年次大会発表論文集, pp.559–562 (2012)
- [8] 大南 勝, 掛谷 英紀: 政治的立場の異なるツイッターユーザー群の特徴分析, 言語処理学会第 22 回年次大会発表論文集, pp.537–540 (2016)
- [9] 工藤 拓, 松本 裕治: チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, Vol.43, No.6, pp.1834–1842 (2002)
- [10] Taku Kudo and Yuji Matsumoto: Japanese Dependency Analysis using Cascaded Chunking, *Proc. of the 6th Conference on Natural Language Learning 2002*, pp.63–69 (2002)
- [11] T. Kudo, K. Yamamoto and Y. Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, *Proc. of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp.230–237 (2004)