

Twitter の多軸的感情情報と 政党支持率との関係について

増井 佑亮 藤野 巖 山本 宙

東海大学大学院情報通信学研究科情報通信学専攻

6bjnm010@mail.u-tokai.ac.jp, fujino@tokai.ac.jp, hiroshi@tokai.ac.jp

1 はじめに

今日の日本の政治情勢を見ると以前にも増して、世論調査の占める地位が高まっているように見える。主要新聞社・放送局は、ほぼ毎月、世論調査を実施しており、その調査結果は定期的に報道されている。各社が必ず行うのが内閣支持率と政党支持率の調査であり選挙に際し政党組織や候補者個人の運動方針決定の一つの重要な指標になっている。その世論調査の手法として個別訪問面接聴取法、RDD 方式、インターネットの三つの手法がある。もっともよく用いられる手法が RDD 方式でこれはコンピュータで乱数計算を基に電話をかけ、応答した相手に質問を行う方式である[1]。

そうした中で近年、ますます多くの有権者がソーシャルメディアを利用するようになったため、候補者も知名度の向上や選挙地盤の強化などを期待して、選挙運動にソーシャルメディアを利用するようになってきた。国内で初めてインターネットの利用が解禁された2013年の参議院議員選挙においても少なくない候補者が選挙期間中にソーシャルメディアを利用し有権者に働きかけていた。ソーシャルメディアにおけるユーザの活動はデータが取得しやすいということもあり、ソーシャルメディアにおけるユーザデータを利用して社会動向の説明や予測を試みる研究が活発に行われてきつつある。

Twitter からのデータを利用して選挙結果の予測を試みる研究もあり実際に2016年に行なわれたアメリカ大統領選では民主党候補のヒラリー・クリントンがアメリカ大統領選に当選すると予測した世論調査が多く見受けられたが、主要ソーシャルメディアを分析する企業の多くは、ドナルド・トランプが選挙で勝つ可能性を予測していた[2]。このように、実用化に向けて多くの国で研究が進められている。後述でも述べるが、その一方でアカウントを登録した人数やツイート数に注目する研究はいくつかあるものの高い精度は得られていないのが現状である。

Twitter は140文字以内短文を投稿する為、他の SNS やブログより手軽にツイートを投稿することが可能

である。Twitter には今起こったことや感じたことについて投稿している場合が多く、その為ツイートにはユーザのその時々感情が素直に現れている場合が多い。その為 Twitter ユーザの発言内容を分析することで、より精度の高い政党支持率の予測を可能にできるのではないかと考えた。ツイートから感情を読み取る研究は実際にいくつか行われている。その中でテキストに表現されている感情をポジティブ・ネガティブの2種類で特徴づけるための感情極性辞書は数多くある。例えば Yahoo! が提供している機能に Yahoo! つぶやき感情分析がある[3]。これは検索窓にキーワードを入力すると、そのキーワードが含まれる話題についてユーザがどのような感情を抱いているかをネガティブ・ポジティブの2極で表現したものである。しかし、政党支持率の予測を行うにあたり重回帰分析上で変数が多い方がよい結果を得られると考えた。しかし、ネガティブ・ポジティブの2極性を利用する研究が多い一方で多種類に分類する研究はあまり行われていない。複数の分類を用いた研究として、中村氏の感情表現辞典[4]では「喜」、「好」、「安」、「哀」、「厭」、「怖」、「怒」、「恥」、「昂」、「驚」の10軸に感情を分類されている。そして、それらの感情軸各々に感情を示す単語が付与されている。この感情表現辞典を参考にツイートから感情情報を数値化できるシステムを作成しツイートの感情値を求め、そこから重回帰分析を用いて政党の支持率を予測する手法を提案する。こうすることによって、ツイート内容にある感情情報を加味することでより高い精度を目指すと共に従来の世論調査と比較することで実用性を示す。

2 従来の研究の流れ、問題点

Twitter を用いた選挙結果の予測を行う先行研究はそのアカウントを登録された人数や発言のお気に入りやリツイート数など有権者に焦点を当てる研究が、多くツイート内容に焦点を当てる研究はあまり多くない。例えば Camero 氏の研究[5]では、フォロワーの数が多ければ候補者が高く認知されており選挙に当選する確率が高いという考えに基づき、候補者のフォロワー数の

推移から選挙当落の予測が試みられたが、実験では高い精度は得られなかった。これは Twitter を利用するクラスタと実際に選挙を投票するクラスタが違うことや汚職や不祥事と言ったマイナスの要因で注目される事の方が政治カテゴリのメディアでは多く一概に注目度が選挙予測に影響するとは考えにくいからである。そこで登録人数や注目度から予測する従来の手法とは異なるアプローチから政党予測を試みることにした。感情分析を利用した選挙予測を行った研究については、Andranik 氏らの研究などが挙げられる[6]。この研究は、ドイツの選挙に関する政党などの選挙関連語を含むツイートの感情分析を行なっている。ツイートの感情分析を行なうにあたり LIWC という心理測定ソフトを利用した。この LIWC は英語の文章を読み取り感情を12次元に分類することができるがドイツ語に対応していない。その為、Andranik 氏らの研究では取得したドイツ語のツイートを自動的に英語翻訳で処理することで、ツイートの感情解析を行なった。そこから得られた情報を元に得られた数値を元に、距離尺度を求めた。この研究では、Twitter と選挙に関する世論調査の結果を相関づけることに成功した。しかしながら日本では、ツイート内容に焦点を当てた政治に関する研究は見受けられなかった。

中村明編「感情表現辞典」を用いた研究では、山本氏らの研究[7]が挙げられる。この研究で感情表現辞典は、10軸の感情のカテゴリがあるが、それを Twitter 上で分類をすることは困難と述べている。この研究は、元々10次元の感情軸がある感情表現辞典を Twitter 向けに次元削除することで改良したものである。手順としては、年齢別に選ばれた被験者にツイートを見てもらい、10軸の感情の中から3つの感情を選ばせ、各感情を10段階に評価する。その結果、特定以上の閾値の感情が複数個ある場合を共起感情と呼び、共起分析やクラスタ分析を用い次元の削除を行った。その結果、「喜」と「好」の感情のペア と「厭」と「怒」の感情のペアは、共起している傾向があり「安」と「昂」の感情は単独で存在することがなく、他の感情を補助する役割があると述べている。しかし、本研究では重回帰分析を用いる分析を行う場合、より多くの説明変数を用いることで高精度な予測結果につながるので感情軸の次元の削除は必要ないと考えた。

3 処理手法

3.1 Twitter の情報を取得と保存

予測モデルの構築にあたり、Twitter streaming data を利用し2014年に投稿されたツイートの中から無作為に抽出された72928717件のデータをデータベースに保存した。取得したデータは CouchDB というデータベースに、アカウント名、ツイート内容、日付や位置情報などの情報と共に保存されている。そのデータベー

スを利用し各政党のキーワードを含むツイートのみを抽出した。

今回解析では自民党と民主党、共産党の三政党を対象に研究を行った。これはデータ数の多い上位三政党であり、他の政党は月ごとのサンプル数が三政党を大きく下回った為、検証対象として信頼性が不十分と考え除外した。

3.2 ツイートから感情情報の抽出

中村明編「感情表現辞典」は、197人の作家の文学作品などを元に文書を分析し、感情に関する単語や慣用句を嬉・怒・哀・怖・恥・好・厭・昂・安・驚の10軸に分類したのである。この感情表現辞書を参考に辞書データを作成しツイートの感情抽出を行なった。具体的には、Twitter から取得したツイートに対して形態素解析エンジンの MeCab を用いて形態素解析を行なう。次に、ツイート内の形態素と感情表現辞書に登録されている感情語の照合を行い合致した感情語をそれぞれの軸に加算することでツイートの感情値を抽出する。例えば、「民主党の政策はすごい」というツイートの場合は、形態素に分けると「民主党 | の | 政策 | は | すごい」となる。ツイート内のそれぞれの形態素と感情表現辞書に登録されている感情語を照合すると、この場合「すごい」が合致する。照合した感情語を各軸にそれぞれをカウントすると、10軸の感情値を算出できる。しかしながら、このままだと表記ゆれが多く発生する為、感情表現辞書と正しく照合しない。例えば感情語辞書に「面白い」という言葉が感情辞書に登録されているが、ツイート内容が「面白かった」だった場合は、感情語辞書では「面白い」と登録されている為合致しない。このような問題を解決するために MeCab には表記揺れを緩和する目的で代表表記という基本語彙が付与されている。そこで、ツイート中の形態素を代表表記に統一し照合をする。この際に抽出されたテキストデータは、計算速度を上げるために、URL などの余分なデータは削除し単語登録した辞書データを使い品詞をカウントする。

3.3 重回帰分析

重回帰分析による支持率の予測を行う。目的変数 y の値を各月の政党別支持率にし、説明変数 x を各月の政党に関するキーワードを含めた。具体的な手順は以下で述べる。本研究では、Twitter と政党支持率の予測を行う。その為に NHK 放送文化研究所[8]の政党支持率のデータを元に2014年の各政党支持率を目的変数に定めた。月ごとの各政党のキーワードを含む感情数から抽出した喜 x_1 、怒 x_2 、哀 x_3 、怖 x_4 、好 x_5 、恥 x_6 、昂 x_7 、安 x_8 、厭 x_9 、驚 x_{10} を説明変数として重回帰分析を行った。下記に重回帰式を示す。

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7 + b_8x_8 + b_9x_9 + b_{10}x_{10}$$

4 検証実験と考察

4.1 Twitterの多軸的感情分析

下記に例として2014年11月に取得したツイート情報をレーダーチャートとしてグラフ化したものを示す。どのグラフも、「嫌い」などの単語が属する厭の軸の割合が最も大きい。これは感情表現辞典に記載されている感情軸ごとの登録単語数に偏りがあるため一概には言えないが政治のトピックには否定的な意見が投稿されやすく、その結果反映されていると考えられる。つまり、その月に起きた話題やニュースによって各月の感情数にばらつきが生じており、これがその時々的情勢に関する民意を反映し政党支持率に影響を与えていると考える。

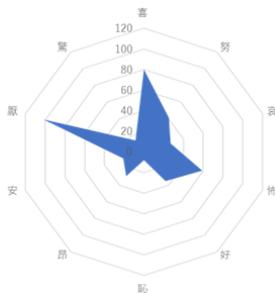


図2 自民党に関する各感情軸のレーダーチャート

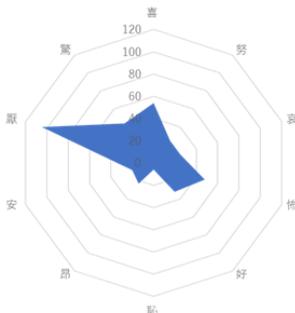


図3 民主党に関する各感情軸のレーダーチャート

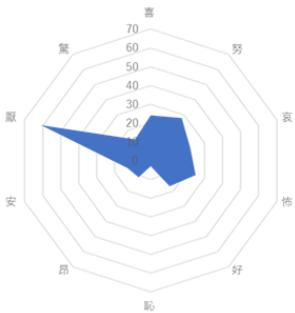


図4 共産党に関する各感情軸のレーダーチャート

4.2 政党支持率との関係

表1は NHK 放送文化研究所の調査による2014年における一カ月ごとの政党支持率である。この政党支持率を目的変数とした。この政党支持率とTwitterの感情解析から得られたデータを元に得られた各政党に関する重回帰式の各係数を表2に示す。

表1: 2014年の政党支持率

月	自民党	民主党	共産党
1月	40.4	5.8	1.6
2月	36.2	5.8	3.3
3月	38.7	6.5	3.3
4月	38.1	7.4	3.6
5月	41.4	5.6	2.4
6月	36.9	5.1	2.8
7月	34.3	4.8	3.4
8月	36.7	6.4	3.2
9月	40.4	5.4	3.3
10月	40.2	5.6	3.3
11月	36.6	7.9	3.5
12月	38.1	11.7	4.3

表2: 各政党の重回帰式の係数

係数	自民党	民主党	共産党
b_0	36.408	11.102	2.500
b_1	-0.008	-0.287	-0.010
b_2	0.308	0.115	-0.015
b_3	-0.029	0.225	0.136
b_4	-0.054	-0.509	0.010
b_5	-0.049	-0.471	0.06
b_6	-1.347	1.176	0.41
b_7	-0.030	0.133	-0.063
b_8	0.541	0.041	-0.78
b_9	0.004	0.299	-0.026
b_{10}	-0.226	0.023	-0.037

定数項が占める割合が高い。これは、政党支持率自体の変化が小さい為だと考えられる。また、恥の軸などのヒット数が少ない軸ほど、偏回帰係数が大きくなる傾向がある。次に、各政党と比較しどの説明変数が最も目的変数を説明しているかを知る為に、標準化偏回帰係数を求めた。その結果を表3に示す。

表3: 各政党の標準化偏回帰係数

係数	自民党	民主党	共産党
b_1	-0.094	-2.756	-0.227
b_2	2.4728	0.759	-0.169
b_3	-0.210	0.791	1.310
b_4	-0.747	-3.491	0.133
b_5	-1.457	-3.314	0.599
b_6	-1.723	1.088	1.287
b_7	-0.202	0.739	-0.628
b_8	1.9174	0.139	-0.549
b_9	0.0910	5.427	-0.809
b_{10}	-1.858	0.167	-0.358

解析により得られた横軸の標準化偏回帰係数は、政党と比べても偏りがなかった。主観的に考えると好印象を持つはずの喜の軸や好の軸の偏回帰係数が政党支

持率にマイナスの影響を与えているという結果が得られ一見、感情を含むツイートと政党支持率の相互関係が曖昧に感じる。これは、今回の実験ではあくまで各政党ツイート中に各政党のツイートが含まれたものを抽出している為、必ずしもその政党を主語にしていな事が一つの原因として挙げられる。例えば実際のツイートでは、野党派が与党派をバッシングしている内容の文章として投稿されているが野党派の文章としてカウントしている為、否定的な感情を示す感情がプラスとして評価されるケースなどである。実際に恥の軸では野党が「恥を知れ」などの過激な言葉で与党の汚職問題などを指摘している場合を確認した。その様なツイートが自民党にマイナスの影響を与え、野党派の民主党や共産党にプラスの影響を与えたと考える。表4に各党の重回帰分析より得られた相関係数と寄与率を示す。

表4:各政党の重相関係数と寄与率

	自民党	民主党	共産党
重相関係数	0.981	0.994	0.901
寄与率	0.963	0.989	0.813

重相関係数は、実測値と理論値がどれだけ一致しているかを示すものである。このことから、政党支持率と非常に高い相関係数があることが分かる。次に寄与率はYの変動のうちの回帰による変動の割合を表しており、どの政党も高い割合で重回帰式によって支持率を説明できている事が分かる。この重回帰式では説明変数の数が多い為、数値が高くなっている可能性はあるが、重相関係数も寄与率も極めて優位な結果を得られた。

先に3.3節で得られた重回帰式から各月の感情値を代入して予測を試みた。その結果を表5に示す。

表5:各政党支持率の予測値

月	自民党	民主党	共産党
1月	39.3	5.3	2.0
2月	36.7	5.7	3.2
3月	38.6	6.3	2.7
4月	38.5	7.3	3.7
5月	41.7	5.8	2.7
6月	36.8	5.3	2.6
7月	34.3	4.9	3.6
8月	36.5	6.6	2.8
9月	40.3	5.3	3.1
10月	40.3	5.4	3.3
11月	36.4	8.0	3.4
12月	38	11.6	4.2

5 まとめ

今回利用したNHK放送文化研究所の調査は、固定電話のみ利用している。その為、Twitterを利用している若年齢層や携帯電話のみを利用する層がかけ離れていると思われる。実際にインターネットを利用した政党支持率の調査ではクラスタが異なっていることが原

因で、RSS方式を利用した大手の調査とかけ離れた結果になる事が多い現状がある。その様な要因もありインターネットを利用した選挙予測の研究分野を困難にしている。そうした中でSNSを利用した選挙予測の研究は、主に登録者数や投稿数などの大きさから類推する研究が多い中で文章データのみを用いて予測を試み、アンケート調査を行なうことなく高い精度の予測数値を得ることができた。

今回の研究では、選挙期間中のデータを取得することができなかったが今後より多くの情報をデータベースに蓄積することで、手法を変更せずとも政党支持率の予測だけでなく選挙の予測も可能になると考える。

また、このシステムは政治の分野に関係なく様々な分野の予測に応用ができる研究と考える。今回は、政治と言うカテゴリの中で予測を試みたが説明変数となる時間軸に沿ったテキストデータと、結果となる月ごとの売り上げなどの定量データさえあれば予測が可能なので、マーケティング戦略への応用、10次元の感情軸という性質上、本や音楽や動画といったメディア媒体の方がより感情は表れやすいので、そういった方面への応用も期待できるだろう。

参考文献

- [1] 世論調査 : <https://ja.wikipedia.org/wiki/%E4%B8%96%E8%AB%96%E8%AA%BF%E6%9F%BB> (2017/01/12 アクセス)
- [2] アメリカ大統領選を予測していたのは、ソーシャルメディア？ 世論調査よりも正確な理由：
http://www.huffingtonpost.jp/techcrunch-japan/social-media-polls_b_12976650.html (2017/01/12 アクセス)
- [3] ヤフー「つぶやき感情分析」正式公開 検索ワードへのTwitterの反応をネガ・ポジ判定：
<http://www.itmedia.co.jp/news/articles/1411/12/news089.html> (2017/01/12 アクセス)
- [4] 中村明:感情表現辞典,東京堂出版,1993年5月
- [5] Michael P. Cameron, Patrick Barrett, Bob Stewardson: Can Social Media Predict Election Results? Evidence from New Zealand (May 2013)
- [6] Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, Isabell M. Welpe: Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment, Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010
- [7] 山本 湧輝, 熊本 忠彦, 灘本 明代:ツイートの感情の関係に基づく Twitter 感情軸の決定, DEIM Forum 2015, F5-2
- [8] 政治意識月例調査:
<http://www.nhk.or.jp/bunken/yoron/political/2014.html> (2017/01/12 アクセス)