

多党制を考慮したマイクロブログ上の政治的分極化における 党派横断的な情報拡散の分析

大前 拓巳[†] 豊田 正史^{††} 張 翔[†] 吉永 直樹^{††}

[†] 東京大学大学院情報理工学系研究科 〒113-8656 東京都文京区本郷 7-3-1

^{††} 東京大学生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

E-mail: [†]{taktak,toyoda,cs,ynaga}@tkl.iis.u-tokyo.ac.jp

あらまし 世論が保守とリベラルとに二極化する政治的分極化は、激しい論争やデモに繋がることがあり、社会に与える影響が大きい。政治的分極化はマイクロブログ上でも観察されており、保守リベラルの二極化を前提として推定された、ユーザのイデオロギーのスコアやラベルに基づく分析が行われている。しかし日本などの多党制を敷く国では、政党ごとに異なるイデオロギーを持つことが予想されるため、既存の推定手法では多党制下でのイデオロギーの多様性が無視されてしまう。そこで本研究では Latent Dirichlet Allocation を用いることで、日本の Twitter ユーザのイデオロギーをベクトルとして推定する手法を提案し、ユーザの多極化したイデオロギーと党派横断的な情報拡散行動の関係を分析する。その結果、今まで保守リベラルの分断として捉えていた分極化が実は特定のイデオロギー間でのみ起きており、その他のイデオロギー間では寧ろ党派横断的な情報拡散が見られることを示す。

キーワード 政治的分極化、マイクロブログ、グラフ分析

1 はじめに

選挙や政策などをめぐる保守派とリベラル派の政治的な意見の対立は、政治家のみならず世論をも巻き込んで激しい論争やデモに繋がることがあり、社会に与える影響も大きい。例えば 2020 年にはタイで大規模な反政府デモが起きており、国際的な問題となっている¹。政治的分極化あるいは分極化と呼ばれるこうした政治的対立は、特にアメリカではこの 20 年間で深刻化していることが報告されており [1]、強い関心を集めている。

分極化は実社会の政治家や有権者 [1-3] のみならず、マイクロブログなどのオンライン上のユーザ [4-6] 間でも観察されることが分かっている。例えば広く普及しているマイクロブログのひとつである Twitter では、保守・リベラルのユーザ間での情報拡散ネットワークの分断が確認されており [4]、ユーザが自身と近い意見しか触れることのないようなエコーチェンバー現象の存在が指摘されている [5,6]。

マイクロブログ上の分極化を分析するためにはユーザの政治的なイデオロギーを推定する必要がある、これまでに保守リベラルの二極化を前提としたスコアリング・ラベリング手法が提案されている [4-10]。こうして推定したイデオロギーに基づき、これまで分極化におけるイデオロギーと実際のユーザ行動との様々な関係性が分析されてきた [5,6,11-13]。

しかし既存の分極化の分析は二大政党制を採用しているアメリカを中心に行われているため、日本のような多党制における分極化を分析する場合、既存のイデオロギー推定手法ではイデオロギーの多様性が無視されてしまう。例えば維新の会は

ニュース等で第三極として取り上げられることも多く²、日本の政治が多極化していることを示唆する政党と言える。日本でも保守リベラルの対立軸こそ一般的ではあるものの、維新の会を保守とみなすかリベラルとみなすかは世代差があることが知られており [14]、アメリカほど統一的なイデオロギーの見解はない。従って、日本のマイクロブログ上のユーザのイデオロギーもいくつかの政党のまとまりに基づいて多極化している可能性がある。またこのとき、注目している政治的な論点もイデオロギーごとに異なり、政治的な話題に応じてユーザの行動傾向が大きく異なることも予想される。

そこで本研究は、3 つ以上の極を持つ政治的なイデオロギーをベクトルとして考慮することで、ユーザのイデオロギーとマイクロブログ上での行動との関係をより精緻に理解することを目指す。具体的にはまずユーザのイデオロギーを推定する手がかりとして、研究室で収集している Twitter データから、ユーザが過去にどの政治家に何回言及したかの情報 (政治家言及ベクトル) を作成する。その後、Latent Dirichlet Allocation (LDA) [15] を用いることで、多党制を考慮した日本の Twitter ユーザのイデオロギーベクトル (多極的イデオロギー) の推定手法を提案し、モデルが適切に多極化したイデオロギーを捉えていることを評価する。最後に多極的イデオロギーを用いて日本の Twitter 上の政治的な話題を分析することで、今まで保守リベラルの分断として捉えていた分極化が実は特定のイデオロギー間でのみ深刻になっていること、その他のイデオロギー間では寧ろ党派横断的な情報拡散が見られることを確認し、提案手法の有効性を示す。

1: <https://www.afpbb.com/articles/-/3317611>

2: <https://www.sankei.com/politics/news/190405/pl1904050047-n1.html>

2 関連研究

マイクロブログ上の政治的分極化は主に二大政党制のアメリカを対象とし、推定されたユーザのイデオロギーのスコアやラベルに基づいて、Twitter を中心に様々な研究が行われている [4-6, 16-18]。本章では、ユーザのイデオロギーの推定手法、およびイデオロギーと情報拡散行動との関係を分析した研究について述べ、その後日本のマイクロブログ上での分極化に関する研究について説明する。

ユーザのイデオロギーを推定する手法は大きく分けてユーザ同士の結びつきを表すネットワーク情報を用いるもの [4, 5, 7] やユーザの投稿内容を用いるもの [6, 8, 9, 13]、あるいはその両方を用いるもの [10] が挙げられる。ネットワークの情報を用いた例として、Barberá ら [5] は、Twitter ユーザがどの政治家をフォローしているかを表す隣接行列を特異値分解することでユーザのイデオロギーの強さを定量化できることを示した。ユーザの投稿内容を用いた例として、Garimella ら [6] はアメリカの主要なニュースドメインの保守リベラル度合いを表すデータを用いて、ニュース URL を含む投稿をした Twitter ユーザのイデオロギーの強さを定量化した。しかしこれらの研究は保守リベラルの二極化を前提としてユーザのイデオロギーをスコアリング・ラベリングしているため、多党制を敷く国における多極化したイデオロギー間の違いを捉えることができない。

推定したイデオロギーを用いて、これまでにマイクロブログ上のユーザのイデオロギーと情報拡散行動にどのような関係が見られるか、様々な研究がなされてきた。特にユーザの保守リベラルの立場やイデオロギーの強さが似たユーザ群に共通する行動の傾向、あるいはイデオロギーが異なるユーザ間での行動の違いを比較分析することで、マイクロブログ上での分極化について理解が深められている。例えば Conover ら [4, 16] は保守派のユーザはリベラル派に比べて Twitter 上で政治的に活発であることを実証すると同時に、政治的に中庸な Twitter ユーザが党派横断的なリツイートを行いやすいことを明らかにした。Barberá ら [5] は、Twitter ユーザのリツイート関係が保守とリベラルに分断する現象が政治的な話題にしか見られない傾向があることを示し、これまでの研究はソーシャルメディア上での保守リベラルの分断の度合いを過大評価している可能性がある」と結論づけた。しかしながら、これらの研究についても二極化を前提としたアメリカでの事例を主に扱っているため、多極化したイデオロギーに応じてユーザの情報拡散行動がどのように異なるかについては未だ明らかになっていない。

このようにアメリカでのマイクロブログ上の分極化に関する研究は数多くあるものの、日本におけるマイクロブログ上の分極化に関する研究は数が少なく [19-22]、ユーザのイデオロギーの定量化についても保守リベラルの二極化を前提とした研究が行われている。三輪 [20] は Barberá ら [5] の手法を用いて推定した日本の政治家 Twitter アカウントのイデオロギーと、政治家調査データから推定したイデオロギーとに高い相関があることを示し、イデオロギーの定量化に関する既存手法が日本の政

治的な Twitter アカウントにも適用可能であることを示した。しかしこれらの研究では政党ごとの多様なイデオロギーを捉えられておらず、また分析の対象も主に政治家公式アカウントが中心であるため、日本のマイクロブログ上のユーザ間の分極化について十分に明らかになっていない。本研究はこの多様性を考慮したうえでユーザのイデオロギーを定量化し、多党制下における分極化についてより深い理解をもたらす。

3 提案手法

本章では、Latent Dirichlet Allocation (LDA) [15] を用いることで、多党制を考慮した日本の Twitter ユーザのイデオロギーの推定手法を提案する。ここで我々は、異なるイデオロギーを持つ人々は政治家への支持傾向も異なるという点に着目した。具体的には日本の Twitter ユーザがどの政治家に言及したかを表す政治家言及ベクトルをクラスタリングすることで、ユーザの多極化したイデオロギーをベクトルで表現すると同時に、どの政党同士がひとつのイデオロギーを形成するかを明らかにする。以下では、提案手法の構成要素について述べる。

3.1 ユーザの支持政党を表す政治家言及ベクトルの作成

Twitter ユーザのイデオロギーを特定する手がかりとして、ユーザが過去にどの政治家に何回言及したかの情報 (政治家言及ベクトル) を用いる。具体的には、あるユーザが日本の政治家アカウントに対して 1 年間のうちに何回メンション・リツイート・引用リツイートを行ったかを個別に集計したベクトルを作成する。すなわち、各ユーザに対して「政治家の数 $\times 3$ 」の長さのベクトルが与えられることになる。メンション・リツイート・引用リツイートの回数を個別に扱ったのは、これらが用いられる文脈はそれぞれ異なると予想されるからである。

ユーザの投稿の収集には、著者らの研究室において 2011 年 3 月より継続的に収集している Twitter のデータセットを用いた。本データセットは、150 万人程度の公開ユーザからタイムラインを継続的に収集したものである。収集対象のユーザは、2011 年 3 月に 30 名程度の著名な日本人ユーザを選択し、それらのユーザがメンションやリツイートをしたユーザをさらに収集対象として順次拡大していった。政治家言及ベクトルはこのデータのうち 2018 年に行われた全投稿を用いて作成した。

また政治家の Twitter アカウントの収集には meyou.jp³ を利用した。meyou.jp は Twitter 上の著名なアカウントを分野別にまとめたサービスであり、その中にある『政治家・議員』の Twitter アカウントまとめ一覧 ページをスクレイピングすることによって所属政党のラベルがついた政治家アカウントリストを収集した。スクレイピングは 2020 年 1 月 14 日に行った。ただし、民進党など、現時点では既に分裂合併によって存在しない政党もラベルに含まれている点には注意が必要である。収集した政治家の数は政党ごとに表 1 に示す。

3: <https://meyou.jp/>

表 1 収集した政治家アカウント全体の人数および 4 章での評価に使用したアカウントの人数.

政党	全体	評価用
自民党	100	21
自由党	11	6
幸福実現党	11	1
公明党	29	8
共産党	6	5
民進党	68	25
無所属	22	8
日本維新の会	22	16
日本のこころ	4	2
大阪維新の会	6	6
立憲民主党	27	11
社会民主党	3	2

3.2 LDA を用いた多極的イデオロギーの作成

多党制を考慮したユーザのイデオロギーを定量化するために、ユーザの政治家言及ベクトルの集合を教師なしクラスタリングに入力してこれらを分類する. 今回は教師なしクラスタリングの手法として LDA を用いた. LDA は文書の確率的生成モデルとして提案された手法であり、文書の潜在的なトピックを得る際に広く用いられる. 本研究では、ユーザを文書、各政治家への言及を単語とみなして、ユーザの政治家言及ベクトルを LDA に入力することで、多極化したイデオロギーを LDA でモデル化することを実現する.

ここで、LDA が出力した各クラスタがもし政治的なイデオロギーと解釈できるならば、各クラスタへの所属確率はユーザのそのイデオロギーへの傾倒度と解釈できる. そこで本稿では、クラスタ数を k にした場合に LDA によって出力されたユーザの各クラスタへの所属確率を表すベクトルを、そのユーザの k 次元の多極的イデオロギーと定義する. また、以降はこの LDA による多極的イデオロギーのベクトルを \mathbf{I} 、ベクトルの各要素を $I_n (n = 1, \dots, k)$ と表記する.

ただしモデルの学習に使用した政治家言及ベクトルはいくつかの前処理を加えたものとなっている. まず学習に使用する政治家言及ベクトルは総和が 10 以上のものに限定した. 更に年間のリツイート総数が 100,000 回を超えるユーザは異常なユーザとして取り除き、ベクトルの各要素は 685 回を上限値として抑えた. また、LDA のハイパーパラメータはクラスタ数 k を除いて全て Gensim [23] のデフォルト値を用いた.

3.3 提案手法の利点

本研究で提案する多極的イデオロギーは、二極化を前提としたイデオロギーの推定手法に比べ、多党制におけるユーザのイデオロギーの多様性をベクトルとして捉えることができる. 日本では保守寄りと見なされる政党であっても政策によって異なる政治的立場を持つことがニュース等で報道されており⁴、マイクロブログ上のユーザのイデオロギーもそれに伴っていくつ

かの政党のまとまりに基づいて多極化している可能性が考えられる. 多極的イデオロギーはユーザのイデオロギーをベクトルとして捉えるため、二極化を前提としてスカラーで定量化する Barberá らの手法 [5] に比べて、政党に基づくイデオロギーの多様性を捨象することなく表現することができる.

また、LDA は特異値分解などと異なり必ずしも軸が直交するとは限らないため、細かなイデオロギーの違いを捉えられることが期待される. 例えば立憲民主党などは自民党に比べるとリベラル政党であると思われるが、リベラル政党の代表であるような共産党とは政治的な立場は異なり、異なるイデオロギーを持つと考えられる. 軸の直交性を保証するような手法では、政治家言及ベクトルの特徴として最も強いであろう保守リベラルの軸に直交する必要があるため、こうした多極化したイデオロギーを捉えることが難しい. LDA はある程度傾向が似た情報も別のクラスタとして出力することができるため、多党制におけるイデオロギーを捉える手法として適していると思われる.

4 評価実験

LDA は教師なしクラスタリングであるため、提案手法によって得られた \mathbf{I} が、政治的な党派とは無関係なものになっており、イデオロギーとして意味をなさない可能性がある. また、モデルのクラスタ数 k はハイパーパラメータであるため、多党制を捉える上での最適な k の値も自明ではない.

そこで本章では、政治家アカウントを一般のユーザと見なしで作った政治家言及ベクトルが、所属政党のラベル付きデータとして活用できる点に着目してモデルの評価を行う. 具体的には、政治家の政治家言及ベクトルを LDA に入力した際、所属政党ラベルをクラスタリングの正解と見なしたときに得られるモデルの予測精度が、 k を増やすにつれてどのように変化するかを分析する. 同時に、各政党の政治家がどのイデオロギーに分類されやすいかを分析することで、各 I_n がどのような党派と対応しているか解釈を試みる.

4.1 モデルの予測精度の変化

まず、政治家アカウントを一般のユーザと見なしで作った政治家言及ベクトルが、所属政党のラベル付きデータとして活用できる点に着目してモデルの評価を行う. 例えばあるイデオロギー I_n が自民党議員への言及を中心に構成されるものだった場合、自民党のラベルを持つ政治家を一般のユーザと見なしで LDA に入力すれば、この政治家の I_n への所属確率は高くなると考えられる. このように、もしクラスタリングによって得た \mathbf{I} がそれぞれイデオロギーと対応する場合、同じ政党の政治家群はひとつの I_n に所属しやすいことが期待される. ここには「ある政党の政治家は同じ政党の別の政治家に言及しやすい」という仮定があるが、このようなホモフィリー性は米国の政治家の間で観測されており [24]、日本の政治家に対して適用しても支障ないと判断した.

以上の仮定のもと、所属政党ラベルをクラスタリングの正解と見なしたときに算出されるモデルの予測精度が、モデルの k

4: <https://www.nikkei.com/article/DGXMZ046823760R00C19A7PP8000>

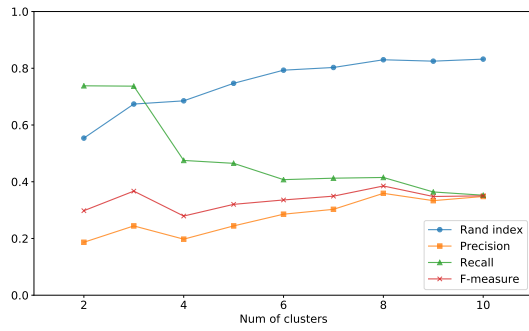


図 1 LDA の k を増やした場合の Rand index, Precision, Recall および F 値の変化.

ラスタ数 k を増やすにつれてどのように変化するかを分析した. 評価に用いる政党ラベル付きの政治家言及ベクトルは, ベクトルの各要素の和が 10 以上のもののみを使用した. ツイートが取得できなかったり, 先述した条件によってフィルタリングされたりした政治家が存在したため, 表 1 に評価に使用した政治家アカウントの数を記載する. 実験では Rand index [25], Precision, Recall および F 値を求めた. Rand index とは正解ラベルがある場合のクラスタリングの評価手法で, N 個のデータから $N(N-1)/2$ 種のペアを選んだときに, 「予測ラベルが一致する」かつ「正解ラベルが一致する」組を真陽とみなすことで Accuracy を求める手法である. Precision と Recall についても同様にして算出した. 結果を図 1 に示す.

特筆すべきは Recall が $k=4$ 以降大きく下がっている点, その一方で Precision や Rand index は $k=8$ まで上がっている点である. Recall は政党ラベルが一致する場合のうち, クラスタリングのラベルと一致する確率を表しているため, Recall が下がるということはひとつの政党内でも I_{n1} と I_{n2} というように異なるイデオロギーに割り当てられた政治家が増えたことを示唆している. 一方で Precision はクラスタリングのラベルが一致する場合のうち政党ラベルが一致する確率を表しているため, Precision が上がるということはひとつの I_n に含まれる政党の数が少なくなっていることを示唆している. これらの結果について, 次節で考察を行う.

4.2 多極的イデオロギーの解釈

本節ではモデルの評価結果の考察として, 多極的イデオロギーの各 I_n がどのような要素を中心に構成されているか, また評価に使用した政治家がどのような I_n に割り当てられているかについて分析する. LDA では各 I_n における各単語の出現確率を求めることができるため, 今回は各 I_n において特定の政治家がリツイートされる確率などを求めることができ, これによって I_n の解釈が可能になる. 更に本実験では政治家を一般ユーザと見なして LDA に入力する評価も行ったため, 特定の政党の政治家が, どの I_n に分類されやすい傾向があるかを確認することによって I_n の解釈ができる. そこでまず $k=3$ と $k=8$ のときに, モデルの各 I_n がどのような要素で構成さ

表 2 $k=3$ (上段) と $k=8$ (下段) における LDA モデルの各 I_n を構成する上位 10 要素. 紙面の都合上, $k=8$ については I_1 から I_3 までの紹介に留める.

I_1	I_2	I_3
志位和夫 (RT)	枝野幸男 (RT)	佐藤正久 (RT)
小池晃 (RT)	蓮舫 (RT)	足立康史 (RT)
日本共産党 (RT)	枝野幸男 (MT)	安倍晋三 (RT)
玉城デニー (RT)	玉木雄一郎 (RT)	世耕弘成 (RT)
有田芳生 (RT)	堀越けいにん (RT)	吉村洋文 (RT)
福島みずほ (RT)	有田芳生 (MT)	杉田水脈 (RT)
小沢一郎 (RT)	有田芳生 (RT)	河野太郎 (RT)
森ゆうこ (RT)	蓮舫 (MT)	松井一郎 (RT)
枝野幸男 (RT)	安倍晋三 (QT)	長尾たかし (RT)
蓮舫 (RT)	杉田水脈 (QT)	足立康史 (MT)

I_1	I_2	I_3
吉村洋文 (RT)	枝野幸男 (RT)	足立康史 (RT)
松井一郎 (RT)	玉木雄一郎 (RT)	杉田水脈 (RT)
吉村洋文 (QT)	原口一博 (RT)	佐藤正久 (RT)
吉村洋文 (MT)	保坂展人 (RT)	安倍晋三 (RT)
大阪維新の会 (RT)	石破茂 (RT)	長尾たかし (RT)
東とおる (RT)	田中康夫 (RT)	吉村洋文 (RT)
浅田均 (RT)	井出ようせい (RT)	丸山穂高 (RT)
松井一郎 (QT)	長島昭久 (RT)	松井一郎 (RT)
佐々木りえ (RT)	津村啓介 (RT)	足立康史 (MT)
松井一郎 (MT)	逢坂誠二 (RT)	日本のこころ (RT)

れているかを表 2 に示す. $k=3$ と $k=8$ との結果を比較すると, $k=3$ の方がモデルの解釈が容易であることが挙げられる. 例えば $k=3$ の I_1 については, 志位和夫や小池晃, 玉城デニーなど共産党や自由党といったリベラル政党へのリツイートを中心に構成されている. 同様に見ると I_2 は民進党や立憲民主党の政治家への言及が主になっており, I_3 は維新の会や自民党などの与党勢力が中心となっている. このようにひとつの政党がひとつの I_n にまとまっている点は, 前項で確認したモデルの Recall が高いという結果と対応していると考えられる. $k=8$ の場合は対照的に, 例えば維新の会の政治家が I_1 と I_3 で混在しているなど, 特定の政党が複数の I_n にまたがっているものが多く見られた. これはモデルのクラスタ数 k を増やすにつれて Recall が下がる結果と対応していると考えられる.

次に, LDA に入力された各政党の政治家が, どの I_n へ所属しやすいかを分析した. 政治家の各 I_n への所属確率について政党ごとに平均をとった結果を表 3 に示す.

$k=3$ と $k=8$ との結果を比較すると, $k=8$ の方はひとつの I_n に含まれやすい政党の数が少ないことが挙げられる. 例えば $k=8$ の I_1 については, 維新の会の政治家の所属確率が極めて高いものの, 他の政党の政治家についてはほとんどゼロに近い確率を示している. 同様に見ると I_4 は民進党と立憲民主党の政治家が, I_8 は共産党と自由党の政治家が高い所属確率を示している. このように $k=8$ のほうがひとつの I_n に含まれる政党の数が少ない点は, 前項で確認したモデルの Precision が高いという結果と対応していると考えられる.

表 3 $k = 3$ (上段) と $k = 8$ (下段) における政治家の各 I_n への所属確率について政党ごとの平均.

	I_1	I_2	I_3
幸福実現党	0.05	0.05	0.90
大阪維新の会	0.00	0.12	0.88
日本維新の会	0.01	0.23	0.75
公明党	0.05	0.20	0.74
自民党	0.04	0.23	0.73
日本のこころ	0.00	0.46	0.54
立憲民主党	0.14	0.84	0.01
民進党	0.11	0.79	0.10
無所属	0.09	0.67	0.23
社民党	0.99	0.01	0.00
共産党	0.73	0.26	0.01
自由党	0.65	0.34	0.00

	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8
大阪維新の会	0.98	0.00	0.00	0.00	0.00	0.01	0.00	0.00
日本維新の会	0.64	0.07	0.10	0.08	0.00	0.10	0.00	0.00
幸福実現党	0.02	0.02	0.64	0.02	0.02	0.24	0.02	0.02
日本のこころ	0.00	0.00	0.50	0.00	0.01	0.49	0.00	0.00
立憲民主党	0.00	0.00	0.00	0.55	0.00	0.01	0.36	0.05
民進党	0.03	0.17	0.03	0.45	0.02	0.09	0.17	0.04
自民党	0.02	0.02	0.10	0.13	0.42	0.28	0.02	0.02
公明党	0.02	0.02	0.02	0.07	0.02	0.80	0.02	0.02
無所属	0.03	0.14	0.03	0.19	0.04	0.38	0.12	0.04
社民党	0.00	0.00	0.00	0.02	0.00	0.00	0.06	0.91
共産党	0.00	0.00	0.00	0.34	0.00	0.00	0.00	0.63
自由党	0.00	0.02	0.00	0.21	0.00	0.04	0.11	0.61

これらの結果から、政治家言及ベクトルを LDA に入力することによって得た多極的イデオロギー I は、多党制におけるユーザの多様なイデオロギーを捉えていると考えられる。また LDA のクラスタ数 k を増やしてもモデルは機能し、より細かい単位での党派を捉えたイデオロギーになることが示唆される。

5 事例分析

本章では多極的イデオロギーの有効性を確認するため、Twitter 上の政治的な話題に関する事例分析を行う。特に本研究ではユーザ間の情報拡散関係を表すリツイートグラフに着目し、提案手法がどのような知見をもたらすのかを明らかにする。以降では、4 章の結果に基づき、 $k = 8$ の LDA モデルによって出力された 8 次元の多極的イデオロギーを用いて分析を行う。

5.1 Twitter 分極化データセットの構築

研究室で収集しているツイート集合から、人手で収集した複数のキーワードを用いて政治的な話題に関するツイートの抽出を行った。表 4 に示される話題の選定・期間・キーワードは Google トレンド⁵などの検索結果を参考に手動で定めた。

話題ごとのツイート集合から、リツイートと引用リツイート

表 4 収集した話題・期間・ツイート数.

話題	期間	ツイート数
森友学園	2017/02~2017/03	2,626,835
朝鮮非核化	2018/04~2018/07	580,585
米大統領選	2016/10~2016/11	701,758
衆院選	2017/09~2017/10	2,461,147
辺野古移設	2018/12~2019/03	1,680,605
イスラム国	2015/01~2015/03	1,156,404
欧州難民	2015/08~2015/11	49,630
豊洲移転	2018/09~2018/10	113,570
桜を見る会	2019/11~2019/12	1,323,681

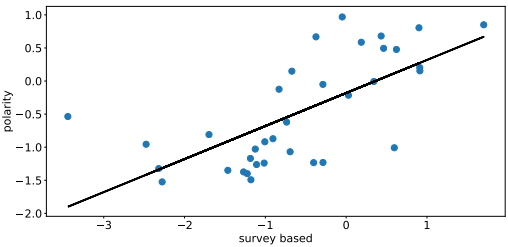


図 2 政治家言及ベクトルおよび政治家へのアンケート調査から推定された保守リベラルのイデオロギーの散布図.

の情報に基づいてユーザ間の情報拡散を表す有向グラフ (リツイートグラフ) G を構築した。あるユーザ x のツイートを別のユーザ y がリツイートまたは引用リツイートしたとき、 x から y の向きにエッジを張ることで G は構築される。このとき G 上のエッジは、ユーザ間の情報拡散の経路と方向とを表していると見なせる。また、政治的な話題についてのリツイートグラフでは、保守リベラル間で分断が起き、両者を繋ぐエッジが少ないことがアメリカでの事例分析から知られている [4]。この知見を活かし、クラスタリングアルゴリズム (METIS [26]) を用いてグラフを 2 つのパーティションに分割した。

更に、本研究で提案した多極的イデオロギーとの比較を行うため、保守リベラルの二極を前提としたイデオロギーのスコアを Barberá ら [5] の研究に基づいて推定した。この手法はユーザと政治家との関係を表す行列を特異値分解した後、第一特異値を用いることでユーザの保守リベラルのイデオロギーを定量化するものである。また同手法は行列の一部を用いて特異値分解をした後、そこで得られた特異値を用いて行列全体のイデオロギーを推定することが可能である。本研究では、4 章でモデルの学習に用いた前処理済みの政治家言及ベクトルのうち、言及の総和の上位 100 ユーザからなる行列から特異値を得た後、前処理を施していない全ユーザに対してイデオロギーの推定を行った。スコアの評価として、政治家へのアンケート調査を元に推定されたイデオロギーとの相関を確認した。これは 2014 年の東大朝日調査データ⁶に含まれる 30 個の争点項目から順序付き項目反応理論モデル⁷を用いて推定された、衆議院議員の

5 : <https://trends.google.co.jp/trends/>

6 : <http://www.masaki.j.u-tokyo.ac.jp/utas/utasindex.html>

7 : <https://www.rdocumentation.org/packages/MCMCpack/versions/1.4-9/topics/MCMCordfactanal>

表 5 リツイートグラフのノード数・エッジ数・および政治家言及ベクトルの総和が 10 以上となったユーザ数.

話題	ノード数	エッジ数	スコア付ユーザ数
衆院選	105,895	1,181,425	21,058
森友学園	71,825	1,151,136	18,098
辺野古移設	59,064	787,989	19,469
朝鮮非核化	51,493	278,803	19,188
米大統領選	87,936	407,456	15,516
イスラム国	86,579	582,189	14,083
欧州難民	12,532	26,147	5,538
豊洲移転	23,522	74,605	4,025
桜を見る会	57,561	662,353	15,882

イデオロギー位置である．表 2 に示した政治家言及ベクトルから推定された政治家のイデオロギーと，政治家調査から推定されたイデオロギーとの散布図 ($n = 35$) では，相関係数が 0.68 と高い値を示し，政治家言及ベクトルから適切に保守リベラルのイデオロギーが推定できていることが確認された．

また，以降は政治に関心の強いユーザを分析するため，政治家言及ベクトルの総和が 10 以上のユーザに着目する．この条件を満たすユーザをスコア付ユーザ数として，リツイートグラフのノード数・エッジ数の情報と併せて表 5 に記載した．

5.2 多極的イデオロギーを考慮したリツイートグラフ上のユーザ分布

提案手法の有効性を確認するため，まず Barberá らの手法で推定した保守リベラルのイデオロギースコアに基づきリツイートグラフ上のノードを色付けし，多極的イデオロギーの各 I_n の所属確率が高いノードのみを図 3 で可視化した．ノードは赤いほどリベラル，青いほど保守であることを意味するが，図 3 では赤と青のノードが偏って分布している．このことから保守リベラルは分断しており，自身と近いイデオロギーのユーザ同士とリツイートし合っていることが観察される．

更に，本稿で提案した多極的イデオロギーを用いることで，こうしたリツイートグラフの分断はユーザの支持政党に応じて異なる傾向を持つことが示唆される．例えば図 3c や図 3e ではそれぞれ I_5 と I_8 への所属確率が 0.1 以上のノードだけが可視化されている．表 3 の結果と併せると，このノードはそれぞれ自民党・共産党を支持するユーザであり，グラフ上に偏って分布していることが分かる．他方，民進党・立憲民主党のイデオロギー I_4 (図 3b) や公明党・無所属のイデオロギー I_6 (図 3d) に所属しやすいユーザはリツイートグラフ上にまばらに分布している．このことから日本での分極化は，特定のイデオロギー間で分断が観察される一方で，その他のイデオロギー間では寧ろ党派横断的な情報拡散が起きていると考えられる．

このことを定量的に確かめるため，ユーザが最も所属しやすいイデオロギー，すなわち所属確率が最大値をとる多極的イデオロギーの要素 I_n を予測ラベル，グラフのパーティションラベルを正解ラベルに見立てた場合に計算される，クラスタリングの精度 (Rand index) を全ての政治的な話題について求めた．

表 6 ユーザが最も所属しやすい I_n を予測ラベル，グラフのパーティションラベルを正解ラベルに見立てた場合の LDA のクラスタリングの精度 (Rand index).

	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8
朝鮮非核化	0.56	0.56	0.91	0.53	0.71	0.57	0.86	0.93
森友学園	0.54	0.74	0.92	0.51	0.76	0.51	0.94	0.98
イスラム国	0.51	0.59	0.81	0.51	0.68	0.54	0.85	0.92
衆院選	0.54	0.76	0.92	0.50	0.67	0.50	0.95	0.97
米大統領選	0.51	0.63	0.78	0.50	0.63	0.52	0.88	0.91
辺野古移設	0.66	0.53	0.96	0.54	0.87	0.59	0.70	0.78
欧州難民危機	0.54	0.53	0.83	0.53	0.65	0.55	0.75	0.76
豊洲移転	0.57	0.51	0.89	0.56	0.76	0.62	0.79	0.91
桜を見る会	0.60	0.64	0.95	0.52	0.84	0.54	0.86	0.92

ある I_n で Rand index が高い場合，そのことはリツイートグラフ上でもユーザが同じパーティションに所属しやすいことを意味しているため，政治的に分断した同質性の高い党派だと考えられる．言い換えると，Rand index が高いイデオロギーを持つユーザはリツイートグラフ上に偏って分布しやすく，逆に低いイデオロギーを持つユーザはグラフ全体に散らばって分布しやすいとみなすことができる，分析結果を表 6 に示す．

表 6 の結果と表 3 に基づく各 I_n の解釈とを照らし合わせると，いずれの政治的な話題でも I_3 (幸福実現党・日本のこころ)， I_5 (自民党)， I_7 (民進党・立憲民主党)， I_8 (共産党・社民党・自由党) のスコアが高くなっていることが分かる．それに対して I_1 (日本維新の会・大阪維新の会)， I_4 (民進党・立憲民主党)， I_6 (公明党・無所属) はいずれもスコアが 0.5 程度という低い値を示している．このことから，イデオロギーによってリツイートグラフ上の分断の傾向は大きく異なると言える．

また， I_4 と I_7 はともに民進党・立憲民主党の支持層であるにもかかわらず，Rand index が全く異なる挙動を示した点は興味深い． I_4 では Rand index が低く，ユーザがグラフ上に散らばって分布しているのに対し， I_7 では Rand index が高く，ユーザがグラフ上に偏って分布している．この理由を調べるため， I_4 と I_7 とを構成する LDA のモデルの主要な要素を表 7 で確認した． I_4 は多様な政党の政治家へのメンションと引用リツイートが中心なのに対し， I_7 は民進党と立憲民主党の政治家に対するリツイートが中心になっている．このことから，民進党や立憲民主党に傾倒しているユーザは，メンションや引用リツイートで積極的に議論するような人と，リツイートだけ行って傍観するような人とに分かれていることが示唆される．

また，政治家に積極的にメンションや引用リツイートを行うユーザは，リツイートグラフ上でも党派横断的に分布している．メンションや引用リツイートを中心に構成されるトピックは I_4 の他にも，無所属と公明党の政治家が中心の I_6 が挙げられる．これらのイデオロギーに所属しやすいユーザは表 6 の結果からリツイートグラフ上に散らばって分布していることが分かる．すなわちこうしたユーザ群は，政治的分極化においてもリツイートや引用リツイートを通じて党派横断的な情報拡散をする人々だと考えられる．

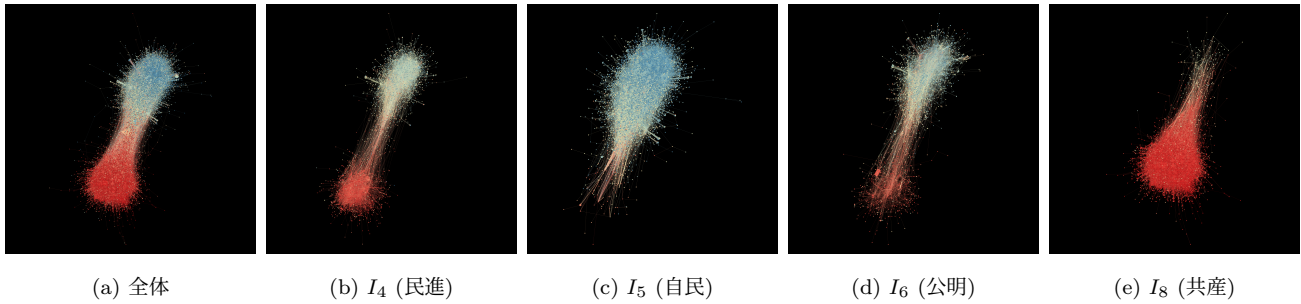


図 3 朝鮮非核化について多極的イデオロギーを持つ全ユーザ (図 3a) と各 I_n の所属確率が 0.1 以上のノードだけを可視化した図 (図 3b~3e).

表 7 $k = 8$ の LDA の I_4 , I_6 , I_7 を構成する要素.

I_4	I_6	I_7
枝野幸男 (MT)	足立康史 (MT)	有田芳生 (RT)
安倍晋三 (MT)	杉田水脈 (QT)	枝野幸男 (RT)
有田芳生 (MT)	金子洋一 (RT)	蓮舫 (RT)
蓮舫 (MT)	足立康史 (QT)	堀越けいにん (RT)
志位和夫 (QT)	あんどろ裕 (RT)	みやけ雪子 (RT)
志位和夫 (MT)	山田太郎 (RT)	玉城デニー (RT)
藤巻健史 (RT)	松浦大悟 (RT)	尾辻かな子 (RT)
小池晃 (QT)	杉田水脈 (MT)	福島みずほ (RT)
小池晃 (MT)	安倍晋三 (QT)	小西ひろゆき (RT)
堀越けいにん (MT)	いさ進一 (RT)	小池晃 (RT)

表 8 朝鮮非核化と辺野古移設での党派横断的なメンションについて、イデオロギーごとのユーザ平均.

		#bridge/all	#mentioned	#mention
朝鮮非核化	I_1	0.163	0.163	0.07
	I_2	0.14	0.267	0.156
	I_3	0.152	0.081	0.606
	I_4	0.206	0.24	0.644
	I_5	0.127	0.118	0.27
	I_6	0.199	0.412	2.038
	I_7	0.132	0.208	0.114
	I_8	0.111	0.159	0.154
辺野古移設	I_1	0.103	0.971	0.529
	I_2	0.222	0.454	0.539
	I_3	0.236	0.276	1.867
	I_4	0.292	0.852	2.246
	I_5	0.184	0.354	0.986
	I_6	0.251	1.154	1.517
	I_7	0.199	0.697	0.704
	I_8	0.183	0.792	1.151

5.3 イデオロギーごとの党派横断的な交流

これまでリツイートグラフ上のユーザの位置について分析したが、次はユーザが実際にどの程度党派横断的な行動をとっているかを理解するための分析を行う。具体的には、各イデオロギーに最も所属しやすいユーザが平均してどの程度党派横断的なリツイート・メンション・引用リツイートを行っているのか調べる。ここで我々は、リツイートグラフのパーティションラベルが異なるユーザ間での交流を党派横断的と定義した。

表 8 では、朝鮮非核化と辺野古移設に関する党派横断的なメンションに限定して、イデオロギーごとのユーザの平均を示している。ここで #bridge/all はあるユーザがその話題に関して投稿した、あるいは受け取った全メンションのうち党派横断的なものの割合、#mentioned と #mention はそれぞれ党派横断的なメンションを送られた数、および送った数の平均を表す。

前節までで I_4 や I_6 のユーザは党派横断的な情報拡散をしやすく、政治的な議論に参加しやすいことが示唆されてきたが、表 8 の #mention の列に注目すると確かにこれらのイデオロギーでは党派横断的なメンションの数が他のイデオロギーに比べて多いことが読み取れる。また本稿に掲載できなかったが、これらのイデオロギーでは党派横断的なリツイートも同様に生じやすいことを確認した。

更に、幸福実現党や日本の心のイデオロギーである I_3 に注目すると、このイデオロギーでも I_4 や I_6 と同様に自ら進んで党派横断的なメンションをしやすことが分かる。この結果は、同じ保守政党とされる自民党のイデオロギー (I_5) と比べると

異なる傾向を示しており、既存手法では捉えられなかったイデオロギーとユーザの行動の関係を表す一例だと言える。

ただし、各イデオロギーの党派横断的な行動の傾向は分析する政治的な話題によって異なる。朝鮮非核化と辺野古移設を比較すると、後者では民進党 (I_7) や共産党 (I_8) のイデオロギーを持つユーザがより党派横断的なメンションを行っている。実はこれらのイデオロギーに含まれる玉城デニーという政治家は、沖縄県知事として辺野古移設の議論の中心にいたため⁸、同じイデオロギーを持つユーザの関心度も高く、政治的な議論を誘発したのだと考えられる。このように話題ごとに異なるユーザの行動をイデオロギーと結びつけて捉えられる点は、提案手法である多極的イデオロギーの有効性を示していると言える。

6 おわりに

本研究では、LDA を用いた多極的イデオロギーを提案し、イデオロギーごとにユーザの性質や情報共有行動の特徴を比較分析することで、日本の Twitter ユーザの多極化したイデオロ

8: <https://www.okinawatimes.co.jp/articles/-/584871>

ギーと党派横断的な情報拡散行動との関係を明らかにした。提案手法の評価では、日本の Twitter ユーザーが複数の異なるイデオロギーに適切に分割されることを明らかにした。更に多極的イデオロギーを用いた事例分析を行った結果、今まで保守リベラルの分断として捉えていた分極化が実は特定のイデオロギー間でのみ深刻になっていること、その他のイデオロギー間では寧ろ党派横断的な情報拡散が見られることを確認し、多党制下での分極化の分析において提案手法が有効であることを示した。

しかしながら、本研究には今後の展望となるようないくつかの制約が存在する。多極的イデオロギーは LDA を用いたシンプルなものであり、モデルを拡張する余地は多いと考えられる。また LDA の学習には 2018 年のデータを用いたが、日本の政治は日々変化しているため、学習に使うデータの時期が異なると、多極的イデオロギーの内実が変わる可能性もある。どの時点のデータでモデルを学習するかは分析結果に少なくない影響を与えると考えられるため、時系列情報が多極的イデオロギーの出力にどのような影響を与えるのかはより詳細な検証が必要である。今後はこうした発展的な課題に取り組み、多党制を採用する国家におけるユーザーのイデオロギーとマイクロブログ上での行動との関係をより精緻に理解することに努めたい。

謝 辞

本研究は、JST、CREST、JPMJCR19A4 の支援を受けたものです。

文 献

- [1] M. P. Fiorina and S. J. Abrams, "Political polarization in the American public," *Annual Review of Political Science*, vol. 11, no. 1, pp. 563–588, 2008.
- [2] P. DiMaggio, J. Evans, and B. Bryson, "Have American's social attitudes become more polarized?," *American journal of Sociology*, vol. 102, no. 3, pp. 690–755, 1996.
- [3] A. I. Abramowitz and K. L. Saunders, "Is polarization a myth?," *The Journal of Politics*, vol. 70, no. 2, pp. 542–555, 2008.
- [4] M. D. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, and A. Flammini, "Political polarization on Twitter," in *5th International AAAI Conference on Weblogs and Social Media*, pp. 89–96, 2011.
- [5] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau, "Tweeting from left to right: Is online political communication more than an echo chamber?," *Psychological Science*, vol. 26, no. 10, pp. 1531–1542, 2015.
- [6] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis, "Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship," in *Proceedings of the World Wide Web Conference*, pp. 913–922, International World Wide Web Conferences Steering Committee, 2018.
- [7] S. Finn, E. Mustafaraj, and P. Metaxas, "The co-retweeted network and its applications for measuring the perceived political polarization," *WEBIST 2014 - Proceedings of the 10th International Conference on Web Information Systems and Technologies*, vol. 1, pp. 276–284, 01 2014.
- [8] A. Rao and N. Spasojevic, "Actionable and political text classification using word embeddings and LSTM," *arXiv preprint arXiv:1607.02501*, 2016.
- [9] A. Bilbao-Jayo and A. Almeida, "Political discourse clas-

- sification in social networks using context sensitive convolutional neural networks," in *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*, pp. 76–85, 2018.
- [10] P. Lahoti, K. Garimella, and A. Gionis, "Joint non-negative matrix factorization for learning ideological leaning on Twitter," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 351–359, 2018.
- [11] U. Chitra and C. Musco, "Analyzing the impact of filter bubbles on social network polarization," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, pp. 115–123, 2020.
- [12] K. Sasahara, W. Chen, H. Peng, G. L. Ciampaglia, A. Flammini, and F. Menczer, "Social influence and unfollowing accelerate the emergence of echo chambers," *Journal of Computational Social Science*, pp. 1–22, 2020.
- [13] J. An, H. Kwak, O. Posegga, and A. Jungherr, "Political discussions in homogeneous and cross-cutting communication spaces," in *Proceedings of the International AAAI Conference on Web and Social Media*, pp. 68–79, 2019.
- [14] M. Endo and W. Jou, "How does age affect perceptions of parties' ideological locations?," *Japanese Journal of Electoral Studies*, vol. 30, no. 1, pp. 96–112, 2014.
- [15] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. null, pp. 993–1022, 2003.
- [16] M. D. Conover, B. Gonçalves, A. Flammini, and F. Menczer, "Partisan asymmetries in online political activity," *EPJ Data Science*, vol. 1, no. 1, p. 6, 2012.
- [17] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on Facebook," *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015.
- [18] L. Lima, J. C. Reis, P. Melo, F. Murai, L. Araujo, P. Vikatos, and F. Benevenuto, "Inside the right-leaning echo chambers: Characterizing gab, an unmoderated social system," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 515–522, IEEE, 2018.
- [19] M. Yoshida and F. Toriumi, "Information diffusion power of political party Twitter accounts during Japan's 2017 election," in *International Conference on Social Informatics*, pp. 334–342, Springer, 2018.
- [20] 三輪洋文, "Twitter データによる日本の政治家・言論人・政党・メディアのイデオロギー位置の推定," *選挙研究*, vol. 33, no. 1, pp. 41–56, 2017.
- [21] 小野塚亮 and 西田亮介, "ソーシャルメディア上の政治家と市民のコミュニケーションは集団分極化を招くのか—Twitter を利用する国会議員のコミュニケーションパターンを事例に," *情報社会学会誌*, vol. 9, no. 1, pp. 27–42, 2014.
- [22] 小林哲郎, "ソーシャルメディアと分断化する社会的リアリティ (<特集> Twitter とソーシャルメディア)," *人工知能*, vol. 27, no. 1, pp. 51–58, 2012.
- [23] R. Řehůřek and P. Sojka, "Gensim—statistical semantics in python," *Retrieved from genism. org*, 2011.
- [24] E. Colleoni, A. Rozza, and A. Arvidsson, "Echo chamber or public sphere? predicting political orientation and measuring political homophily in Twitter using big data," *Journal of communication*, vol. 64, no. 2, pp. 317–332, 2014.
- [25] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical association*, vol. 66, no. 336, pp. 846–850, 1971.
- [26] G. Karypis and V. Kumar, "A fast and high quality multi-level scheme for partitioning irregular graphs," *SIAM Journal on scientific Computing*, vol. 20, no. 1, pp. 359–392, 1998.