

マイクロブログユーザのイデオロギー変容の検知とその要因の分析

張 翔† 豊田 正史†† 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: †{cs,toyoda,ynaga}@tkl.iis.u-tokyo.ac.jp

あらまし アメリカでは保守派のデモ隊が議会に侵入し、タイでは反政府デモが行われるなど、政治イデオロギーの分断が全世界で大きな社会問題となっている。この分断を解消するため、多様なイデオロギーの意見を提示するなどの介入が試みられているが、逆に反発が生じるだけであることが分かっており、情報の適切な提供方法は未だ明らかになっていない。そこで本研究では、Twitter における政治家の投稿を用いてトピックに依存してイデオロギーを変更させるユーザを特定し、その変容の要因を探ることで、異なるイデオロギーに対する態度の軟化に寄与するような情報の提供方法にはどのような要素が有用と考えられるかを探る。

キーワード Twitter, 分断, 計算社会学, イデオロギー

1 はじめに

人は自分の聞きたい意見のみを聞く。現代ではインターネットやソーシャルネットワークサービスの普及により、理論上は関心のあるトピックについて多様な観点から見た情報が得られるようになった。しかしながら、技術的には多様な情報が得られるとしても、実際には人は自分のあらかじめ抱いている価値観や信念に合致する情報のみを取得する。これは、人間には自分の価値観や信念に合致しない情報を無意識的に回避してしまう特性があるためであり、結果として多様な情報が得られないことがままある。

人が自分の聞きたい情報ばかりを聞くようになると、ひいては大規模な社会不安を招く行動に結びつく。例えば、2020年のアメリカでは右翼の過激派によって議会が襲撃される事件が発生したが、複数のメディアによってこの事件が当時の大統領であったトランプ氏の発信する情報のみを信じ、トランプ氏の「扇動」ともとれる投稿を見た支持者らが引き起こしたものである可能性が指摘^{1,2}されている。また日本でもワクチンが陰謀であるという情報以外を聞き入れない過激派の反ワクチン派によってデモ³が引き起こされた。このように、自分の価値観に合致する情報だけを取得することは自分の信念の増幅に繋がり、過激な行動をも肯定する動力源となりうるため、できるだけ多様な情報を取得することが理想的な情報取得のあり方といえる。

人々ができるだけ多様な情報を接種するように働きかけた研究は多く存在するが決定的な手法は未だ存在していない。例えば Gillani らは特定のトピックについて Twitter 全体の情報伝搬を可視化するツールを作成し、このツールの使用者が「自身がいかに部分的な情報しか接種していなかったか」を自認させ自省を促すことを狙った [1]。彼らは実際に Twitter ユーザの

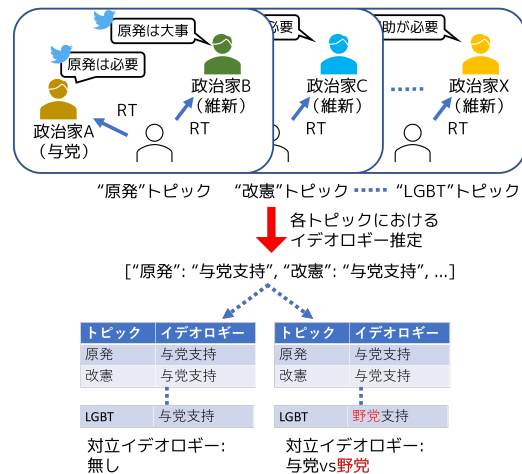


図. 1: 政治家へのリアクションデータを用いたイデオロギー変容ユーザの特定フロー (概要)。

中から被験者を募って実験を行い短期的な有効性は確認したが、長期的な有効性は確認できなかった。他にも Spitz ら [2] は、単純にあるターゲットユーザに多様な思想を推薦するだけでは心理的抵抗が生じるため、そのユーザが尊敬・信頼するインフルエンサーの思想として多様な思想を推薦するというアプローチによって抵抗を弱められると考えた。残念ながら、この手法でも多様な意見への抵抗感を弱めることはできず、それどころか対象ユーザがもともと尊敬・信頼していたインフルエンサーへの信頼感を損ねてしまう傾向が確認されるという結果になった。このように、人々に多様な意見を持ってもらうという目的のために様々な介入が考案されているが、いまだ決定的なアプローチは見つけられていないのが現状である。

そこで今回、そもそもどのような条件下で人は自分の思想とは異なる思想を受容するのかを、マイクロブログサービスのユーザを分析することによって探る手法を考案した。本手法の概要図を図. 1 に示す。本手法はまず前処理としてマイクロブ

1 : www.itmedia.co.jp/news/articles/2101/13/news082.html

2 : www.bbc.com/japanese/57993507

3 : news.yahoo.co.jp/articles/522444c3dd63d59bb8d0833a95720fc5e75ce623

ログデータのトピックによる分割を行う（同図上部）．その後、分析のためにマイクロブログユーザのイデオロギーを政治家へのリアクションデータに基づいて推定する（同図中部）．最後に、推定したイデオロギーをもとに対立するイデオロギーを保持するユーザを見つけ出し、分析する（同図下部）．

分析では、国内最大のマイクロブログサービスといえる Twitter を対象とする．このイデオロギー推定手法によって Twitter 上のトピックごとに各ユーザのイデオロギーを推定し、あるトピックでは与党支持、別のトピックでは野党支持というような両立が起きにくいと考えられる、対立するイデオロギーを同時に抱いているユーザを特定する．更に、特定したユーザの投稿内容やリアクション内容をも分析することで、そのユーザに対立するイデオロギーを抱かせるに至った要因を特定する．Twitter は日本国内では最も使用されているマイクロブログサービスであり、また諸外国でも最も使用されているサービスの一つである⁴ことから実験に用いるサービスとしては適切といえる．

2 関連研究

2.1 他プラットフォームでの先行研究

他者の意見を変更させるような情報提供の仕方については特に Reddit の Change My View（以下、CMV とも）という Sub-Reddit を対象とした研究が行われている．Reddit とは最大手の掲示板サービスの一つであり、より小さな Sub-Reddit というグループに分化する．Change My View とは Sub-Reddit の一つで、Reddit ユーザが任意のトピックについて持論を主張し、他者から意見を募り議論を行うことを趣旨とする．特徴的な機能としては、集まったユーザらと議論を交わし、その結果として持論が変化すると元の主張者が感じた場合には最も持論の変化に影響を与えた相手に Δ という称号を与える機能が存在する．例えば Zeng らはあるトピックの議論に参加したユーザの中から、この Δ を獲得したユーザと獲得しなかったユーザとを対象に発言内容や話法について比較し、元の主張を変更させるユーザに特有の話法などを追及した [3]．

CMV を対象にした研究は一定の成果を示しているが、対象とするユーザが母集団を代表しているかという点については検証の余地がある．すなわち、CMV に参加するユーザはそもそも既に自分以外の意見を聞きに来ている人物であるが、一般的なユーザの多くは「自分以外の意見を積極的に取り入れようとはしない」ユーザであると考えられる．本稿ではそのような一般的なユーザを対象として「どのような方法で情報提供を行えば抵抗なく受け入れてもらえるのか」を考えており、CMV を対象とした既存研究とは相補的な立ち位置になる．

2.2 政治的イデオロギーの定量化

SNS ユーザの政治的イデオロギーを取り扱いやすい形に定量化する手法には先行研究をもとに筆者らが以前に開発した

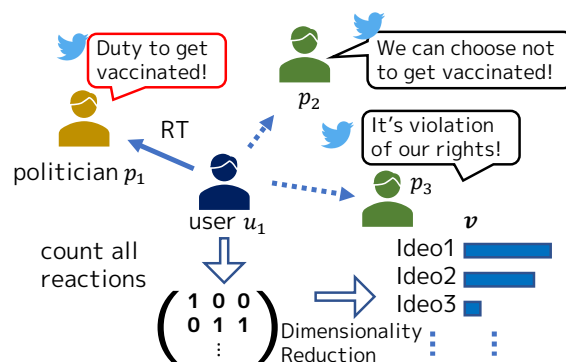


図. 2: 政治的イデオロギー定量化手法の概要図.

手法 [4] を用いる．この手法の概要図を図. 2 に示す．元々は Barberá らが Twitter におけるリツイートのように、著名な政治家へのリアクション（同図上部）を行列の形に直して特異値分解を行い（同図左下）、最大の特異値をもとに各ユーザに割り振られる値をそのユーザの政治イデオロギーとして扱う（同図右下）手法を開発した [5]．この手法は実際のユーザの政治嗜好（支持政党など）にかなり合致することが確認されたものの、アメリカのような二大政党制にしか適用できず日本のような多党制を敷く政治形態の国の分析には不十分であることが指摘されていた [1, 6]．特に、通常の特異値分解ではデータの分散が最大となる軸を新たな次元として使用するため、細かい差異が無視されてしまうという問題点があった．

そこで筆者らは行列分解にテキスト処理の分野で用いられる LDA（Latent Dirichlet Allocation）モデリングを適用することで、多党制にも対応した定量化手法を開発した．LDA は文書のトピック分類に使用される技術で、特定の単語群間の共起パターンを推定する技術であり、例えば買い物客が一緒に買う物など共起パターンが存在すれば広く適用可能である．また LDA は特異値分解とは全く異なったメカニズムで働くため、例えば 2 人のユーザがリアクションを行った政治家が 1 人を除いて全て同じであったとしても、その一人の違いによってこの 2 人のユーザを別々のイデオロギーに割り振ることが可能となる．この手法により、マイクロブログ上の政治家アカウントからの投稿へのリアクションをもとにマイクロブログユーザのイデオロギーをベクトルの形で定量化することが可能となる．

特にこの手法の強みとしてベクトルの次元数を任意に設定することで、従来の保守対リベラルのようなシンプルな二極のイデオロギーだけでなく、維新系列の政党のような、いわゆる「第 3 極」のイデオロギーをも取り扱うことが可能になる．この特徴により、この定量化手法は日本のような多党制政治を敷いている政治状況の分析により適したものとなっている．この定量化手法は政治家の所属政党データを用いた評価実験において、政治家の所属政党と一定の一致度を見せることを確認しており、本稿の日本の政治システムの分という目的に従来の定量化手法よりもより適していると考えられる．

4：中国など一部の国では Twitter は使用が制限されているが Weibo のような類似サービスは存在すると考えられるため本手法の応用性は損なわれまいと考える．

表 1: Twitter データセットの統計.

種類	投稿件数
全投稿 (2018)	11,212,840,419
日本語の投稿	2,049,498,199
政治家の投稿	131,138
政治家の投稿 (公人としての投稿)	80,372
政治家の投稿へのリアクション	3,346,136
- リツイート	2,880,362
- リプライ	232,965
- 引用リツイート	232,809

表 2: 収集した政治家アカウントの人数.

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

2.3 短文用トピックモデル

本手法ではまず、トピックごとに人々が保持するイデオロギーを分析するため、マイクロブログデータをトピックごとに分割するが、マイクロブログデータのような短文の文書に LDA のような通常のトピックモデルを用いることは適切ではない。文書データのトピックによる分割には一般的にトピックモデル技術、特に LDA (Latent Topic Model) モデリングが用いられるが、LDA は文書内の単語と単語との共起関係を利用することから、マイクロブログの投稿や図表のキャプションなどの短文には通常の LDA のパフォーマンスが低下することが報告されている [7, 8].

そこで本稿では、短文のトピックモデリングにより適しているとされる Dirichlet Multinomial Mixture model (DMM) というモデルを用いる。DMM は「マイクロブログのような短文であれば単一のトピックのみを有する」という仮定のもとで LDA の枠組みに則ってトピックを推定する手法 [9] であり、その他の短文用トピックモデルと比較しても特にマイクロブログのようなウェブドメインの短文において、より高い性能を達成するとされる [10]. 本稿では一般的な前処理を施したのち、この DMM モデルを用いて政治家の投稿をトピックごとに分割する。

3 政治的な投稿のデータセット

本節では、分析で使用するデータセットである、マイクロブログ上で政治家アカウントの投稿した政治的な投稿、およびそれらの投稿に対するリアクションデータから構成されるデータセットの構築手順を説明する。

3.1 使用データ

本研究では筆者の所属研究室で 2011 年以降継続して収集している Twitter データセットを使用した。これは最初に 30 名ほどの著名な人物をシードとして設定し、シードの人物の投稿を収集するとともに、シードの人物がリアクションを行った先のユーザを新たに収集対象として再帰的に加えるという手法により構築したデータセットである。使用したツイートの件数を表 1 に示す。

また本研究で対象とする政治家のアカウントは meyou.jp⁵ から収集した。このサイトは著名な政治家数百名についてその Twitter アカウントを (当時の) 政党ごとにまとめているサイトであり、このサイトからスクレイピングを行うことで日本の主要な政治家の所属政党と Twitter アカウントのデータを獲得した。このスクレイピングは 2020 年の 1 月 14 日に行ったため、現在の所属政党とは異なっている点に注意が必要である。各政党ごとに収集できた政治家アカウントの数を表 2 に示す。

3.2 非政治的な投稿の除去

ここで、政治家の投稿の全てが政治的とは限らないため、機械学習によって非政治的な投稿のみを選別して取り除く。例えば新年や四季折々の挨拶、流行りの映画や本などの感想、単なる日常生活に関する投稿などが考えられる。こういった投稿に対してリアクションを起こしたユーザがいたとしても、それが何らかの政治イデオロギーを示唆するとは考えにくい。そこで、簡単なアノテーションを行い、政治的・非政治的な投稿を一定数用意したうえでそれらを識別可能な識別器を用意し、それによって非政治的な投稿のみを取り除いた。

アノテーションの対象として政治家が 2018 年に行った全投稿の中から 388 件ランダムに抽出し、「公人としての投稿といえるかどうか」という観点で筆者が投稿をアノテーションを行った。表 3 にアノテーションされたツイートの例を示す。この例に示すように、単に政治的なトピックのツイートであるかどうかというだけでなく、政治家としての知名度を活かした災害情報の発信や、あるいは経済問題や外交問題に対するスタンスを表明しただけの投稿であっても公人としての投稿とみなした。ただし、今は政治家自身から発される情報に焦点を当てているため、この 388 件にリツイートは含まれないようにした。また、アノテーションの際には投稿の本文中に他サイトへのリンクが含まれていた場合にはそのリンク先も判断材料とした。同様に、アノテーション対象の投稿がリプライであった場合には、そのリプライが含まれる一連の応答を全て参照し、引用リツイート

5: <https://meyou.jp>

表 3: 公人としての投稿であるか否かのアノテーションの例. (URL などは適宜トークンに置換)

公人・私人	投稿内容
公人	行政が、原発を推進し、司法がそれを追認してきた。その結果、福島東電原発事故が起きた。行政や司法の反省はどこにあるのか。
公人	中国の影響力拡大の裏返しだが、トランプ政権による安易な覇権放棄 (abdication) であることも鋭く指摘していましたね。まずは日本の役割が重要になる 2018 年ですね。[URL]
公人	今日の大阪は雲。少し雨も。午前中の様々な会議を終え、午後心は晴れやかに対話を。
公人	台風 24 号に被害が発生しています。どうかお気をつけください。
私人	「エポック」のカレー美味しい！今治市玉川町龍岡の一見普通の民家ですが、入るとアジアンテイスト満載の不思議空間。ぜひ！
私人	おはようございます。巨大な虹が出てました。「君の名は」で隕石が落ちる瞬間に綺麗な虹みたいな映像が夜空に浮かんでました。そんな感じでした。幸せな気分ですね。#君の名は [URL]
私人	おはようございます！今日もがんばっぺ！
私人	箱根駅伝 ゴールした瞬間に選手たちがお辞儀をする大学がある。これも素晴らしい指導。
私人	@[username] ありがとうございます 今気がつきました！読んでいただき感謝です

表 4: DMM 訓練時のストップワード.

事, 今, 何, 問題, 氏, こと, の, よう, さん, ため, これ, もの, それ, 日, 人, 年, 日本, 私, 方, 中, 時, 等, 的, 者, 月, 回

の場合も引用先の投稿も判断材料とした。結果、公人としての投稿としてアノテーションされた投稿は 333 件で、私人としての投稿は 40 件、リンク先の記事がすでに削除されており判定できなかった投稿が 15 件存在した。この内判定できなかった 15 件の投稿を取り除いた 373 件の投稿をデータセットとして分類器を訓練する。

このアノテーションされたデータセットを用いて識別機を訓練し、実際の政治家投稿データセットに適用した。この際、私人としての投稿が公人としての投稿と比べて非常に少ないことから、訓練がうまくいかないことが予測される。そこで、公人としての投稿と同じ件数になるまで、私人としての投稿をランダムにオーバーサンプリングする。特徴量としては各ツイートに対して形態素解析を行うことで名詞だけを抽出し、それらの名詞を埋め込んだ 200 次元のベクトルに対して平均をとったベクトルを使用した。識別機には scikit-learn ライブラリのランダムフォレストモデルを使用し、n_estimators を 90,100,110 から、criterion を gini, entropy から 5-fold cross validation でグリッドサーチを行った。この結果の f1-macro 値は 0.69 であった。この分類器を用いて政治家の投稿を分類し、私人としての投稿を取り除いたあとの政治家の全投稿に対する Twitter ユーザのリアクションデータセットを改めて構築した。このデータセットの詳細を表 1 に示す。以降ではこの私人としての投稿を取り除いたあとの政治家の全投稿を使用する。

4 マイクロブログユーザのトピックごとのイデオロギー推定

本節では、提案手法であるマイクロブログデータの中からトピックごとに異なるイデオロギーを有するユーザを特定する手法について説明する。

4.1 Twitter データセットのトピックごとの分割

まず本稿ではトピックごとのイデオロギー変容を分析対象とするため、全政治家の投稿をトピックごとに分割する。分割には関連研究の節で述べたように DMM モデル [9] を用いる、

DMM 訓練用のコーパスとして、政治家の投稿を形態素解析し Bag-of-Words 形式に直した。ただし今回は形態素解析の後で名詞だけを抽出し、さらにストップワードを取り除く処理を加えた。ここで名詞だけを使用したのは結果の解釈をしやすくするためである。また、ここで使用したストップワードは、一度何もストップワードを使用しない状態で得られた各トピックの単語出現確率の上位語を確認し「トピックの推定にほとんど全く寄与しない」と考えられる語を筆者が選択した。ここでのストップワードを表 4 に示す。こうして政治家の投稿の DMM 訓練用コーパスを構築した。

構築したコーパスを用いて DMM を訓練し、実際に政治家の投稿をトピックごとに分割した。実装は Qiang らの STTM ライブラリ⁶を使用し、デフォルトの設定である 200 イテレーションの訓練を行った後に各投稿に対して各トピックへの所属確率を推定した。Qiang らによると、200 イテレーションあれば十分損失は収束する [10] とされている。この所属確率が最も高いトピックを各ツイートが所属するトピックと仮定し、データセットをトピックごとに分割した。例えば、ある投稿が 80% の確率で「外交」トピックに属し、20% の確率で「経済」トピックに属すると推定された場合、この投稿は「外交」トピックに属するものとして扱う。ここで、ハイパーパラメータであるトピック数は暫定的に 14 としたが、理想的にはトピックのアノテーションがなされたデータセットなどを用いて最適なトピック数のチューニングを行うことが望ましい。

各トピックの単語出現分布の上位 10 語と、そのトピックに属する関連投稿の総数を表 5 に示す。ここで、関連投稿とは、各政治家の投稿のうちそのトピックに属するもの、およびそれらの投稿に対する一般ユーザのリアクションを指す。また、各トピックのラベル名はこれらの語や実際に各トピックに属する

6: <https://github.com/qiang2100/STTM>

表 5: 各トピックの出現頻度上位 10 語と関連投稿の総数.

通知	原発	LGBT	沖縄	与党批判	外交	改憲	維新	経済	国会	森友	行事	災害	支援
ご覧	原発	議員	沖縄	会	米	野党	活動	円	法案	文書	会	被害	子ども
更新	大阪	ん	候補	公明党	北朝鮮	政治	会	%	国会	森友	今日	復旧	必要
出演	視察	弁護士	選	本部	大統領	自民党	維新	日銀	質問	財務省	参加	災害	制度
分	世界	女性	選挙	開催	会談	国民	参加	万	野党	国会	地元	対策	万
ブログ	ニュース	デマ	知事	議員	トランプ	議員	議員	億	審議	改ざん	開催	台風	化
アメブロ	メディア	ツイート	投票	推進	朝	憲法	報告	兆	本会議	答弁	出席	支援	国
公式ブログ	LINE	事件	辺野古	会長	韓国	党	今日	政府	質疑	野党	大会	情報	円
お知らせ	万	ネット	応援	自民党	中国	国会	皆様	国債	衆議院	総理	本日	地震	%
生中継	化	自分	基地	党	首脳会談	代表	開催	国民	参議院	学園	今年	豪雨	歳
知事	万博	批判	当選	要望	米国	維新	本日	国	時間	首相	記念	号	支援
85,969	111,490	176,784	229,639	41,158	281,404	284,982	72,037	123,383	298,209	541,772	87,245	263,722	152,856

投稿を参照し、筆者が付記した。

4.2 イデオロギー推定モデルの訓練

これまでに構築したトピックごとの政治家アカウントの投稿、およびそれらの投稿へのリアクションからなるデータセットを用いて、各トピックにおけるマイクロブログユーザの政治的イデオロギー推定モデルを訓練し、実際に各ユーザの政治的イデオロギーを推定する。基本的な流れは筆者らの以前の研究 [4] と同じであるため詳しくはそちらを参照されたい。

まずトピックごとに分割する前の、私人としての投稿を全て取り除いたあとの全投稿・投稿へのリアクションを用いて各ユーザごとに「どの政治家に何回リアクションをしたか」をカウントする。ここで、本稿でのリアクションとは「リツイート・メンション・引用リツイート」のいずれか（以降それぞれ RT, MT, QT）とし、各政治家・各リアクションタイプごとに別々にカウントする。本稿では 306 名の政治家のアカウントを収集したので、ここで各ユーザに対して $306 * 3 = 918$ 次元のベクトルが割り振られる。

このベクトルを用いて LDA モデルを訓練することで、イデオロギー推定モデルを得る。すなわち、各ユーザを文書、各ベクトルをその文書内に出てくる単語の出現回数として捉え、それらの文書（ユーザ）の各トピック（イデオロギー）ごとへの所属確率を推定するモデルを訓練する。LDA モデルの実装としては先ほどと同じく Qiang らの STTM ライブラリを使用し、イテレーション回数などのハイパーパラメータもデフォルトの値を使用した。このようにして訓練したモデルを使用して獲得した政治的イデオロギーベクトルは、政治家の所属政党データと比較するとある程度一致しており、一定の妥当性を有することを確認済みである。

このイデオロギー推定モデルに対してトピックごとに分割した政治家の投稿へのリアクション回数ベクトルを与えることで、各ユーザの「各トピックにおける」イデオロギーベクトルを獲得する。このイデオロギーベクトルは任意の次元数を設定可能であり、各次元の値は各イデオロギーへの所属確率を意味する。以降の分析では、このベクトルのうち、最も所属確率が大きいイデオロギーを各ユーザのトピックにおける所属イデオロギーとして扱う。例えば、あるユーザに与えられたイデオロギーベ

クトルが $v = (0.1, 0.7, 0.2, 0.0)$ であった場合、最大の値の次元は 1 番目であるためそのユーザは 1 番目のイデオロギーに属するものとして扱う（どのイデオロギーがどのような特徴を持つのかは後述する）。ただし、リアクションの回数が少なすぎる場合にはイデオロギー推定の精度が悪化することが予想されるため、各トピック中で最低でも 10 回はいずれかの政治家にリアクションをとっている場合にのみイデオロギーを推定するものとし、そうでない場合はそのユーザはそのトピックについてイデオロギーを保持しない（興味がない）ものとして扱う。

本稿ではイデオロギー数を 7 に設定し、このときの各イデオロギーに対応するリアクション（リアクション先の政治家とリアクションタイプ）を表 6 に示す。この表から、例えば、佐藤正久議員（自民党）に対して RT を頻繁に行うユーザが存在する場合、そのユーザは与党イデオロギーに属することが読み取れる。ただし、イデオロギーの名前は所属する政治家やリアクションを見て筆者が付記した。この表から、大別すると与党寄り、野党寄り、維新寄りのイデオロギー、および MT/QT を中心とする MT イデオロギーが抽出されたことが読み取れる。この MT イデオロギーを有するユーザは頻繁に MT か QT を行っており、その内容は主に賛同か批判である。更に細かく見ると、野党寄りのグループも自由党・共産党・社民党グループ、共産党単独のグループ、そして立憲民主党単独のグループの 3 つに細分化される。以降では、これらのイデオロギーをそれぞれ与党イデオロギー、野党イデオロギー、維新イデオロギー、MT イデオロギーと呼び、野党イデオロギーを細かく見る場合はそれぞれ自共社イデオロギー、共産イデオロギー、そして立憲イデオロギーと呼ぶ。こうして得たユーザごと・トピックごとの所属イデオロギーを用いて、トピックごとに異なるイデオロギーを有するユーザを特定する。

4.3 トピックごとに異なるイデオロギーを有するユーザの特定

まず、本稿で対象とするユーザの定義を行う。前節までの処理により、各ユーザは 14 個のトピックのそれぞれについて表 6 のいずれかのイデオロギーを有するか、あるいはリアクション回数が最低回数の 10 回に満たないためにそもそもイデオロギーを有さない。ここで、14 個のトピックのそれぞれに対して所持するイデオロギーの種類数によってユーザを以下のように分

表 6: 各イデオロギーで行われる確率の大きいリアクション上位 5 件 (イデオロギー数 7 の場合) .

野党 (自共社)	MT (to 野党)	維新	MT (to 与党・維新)	与党	野党 (共産)	野党 (立憲)
R 小沢一郎 自由 Q 志位和夫 共産 R 足立康史 日維 Q 吉村洋文 大維 R 佐藤正久 自民 R 志位和夫 共産 R 有田芳生 民進						
R 志位和夫 共産 M 志位和夫 共産 R 吉村洋文 大維 Q 吉村洋文 日維 R 安倍晋三 自民 R 日本共産党 共産 R 枝野幸男 立憲						
R 福島みずほ 社民 Q 小池晃 共産 R 吉村洋文 日維 M 足立康史 日維 R 河野太郎 自民 R 小池晃 共産 R 蓮舫 民進						
R 森ゆうこ 自由 M 小池晃 共産 R 杉田水脈 日心 Q 杉田水脈 日心 R 世耕弘成 自民 R 吉良よし子 共産 R 三宅雪子 無所						
R 山本太郎 自由 M 枝野幸男 立憲 R 松井一郎 日維 Q 足立康史 日維 R 杉田水脈 日心 R 玉城デニー 自由 R 尾辻かな子 民進						

表 7: n-ideology ユーザの数.

n	興味を持ったトピック数	人数
Uni	1	799
Uni	2-14	2,332
Bi	2-14	858
Tri	3-14	164
Quadri	4-14	3

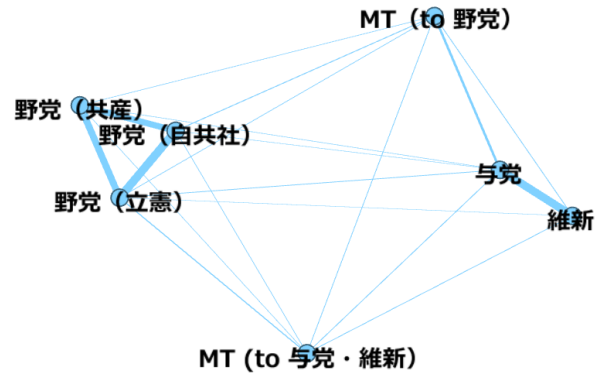


図. 3: イデオロギー間の共起頻度グラフ.

ける.

Uni-ideology ユーザ 所持するイデオロギー数が 1 のユーザ.

Bi-ideology ユーザ 所持するイデオロギー数が 2 のユーザ.

Tri-ideology ユーザ 所持するイデオロギー数が 3 のユーザ.

Quadri-ideology ユーザ 所持するイデオロギー数が 4 のユーザ.

例えばあるユーザが「経済」トピックと「原発」トピックにおいて与党イデオロギー、「改憲」トピックにおいて維新支持, それ以外のトピックには 10 回未満しかリアクションをしておらずイデオロギーを所持していなかった場合, そのユーザは Bi-ideology ユーザとなる.

n-ideology ユーザの数を表 7 に示す. ただし, あるユーザが関心を持つトピックが 1 つのみであった場合 (最低回数以上リアクションを起こしたトピックが 1 つのみであった場合), 必然的にそのユーザは Uni-ideology ユーザ (以降, 単に Uni ユーザ) となる. しかし, 今回は 2 つ以上のトピックに対してイデオロギーを有していた場合に, 各ユーザが意図的に単一のイデオロギーしか有しているケースと区別したため, そのようなユーザはデータセットから取り除く (今回はそのようなユーザは 799 人存在した). また, 所持するイデオロギー数が 5 以上のユーザはデータセット中には存在しなかった.

これまでの処理により, マイクロブログユーザのトピックごとのイデオロギーの推定が可能になり, また実際にトピックごとの政治的イデオロギーのデータセットが構築できた. 更に, ユーザごとに所持する政治的イデオロギーの種類数も付与できた.

5 共起イデオロギーの分析

本節では共起するイデオロギーの分析を行う. 特に, 与党イデオロギーと野党イデオロギーのような, 一般的に共起しにくいことが予想される対立しているイデオロギーペアの共起に着目し, 対立イデオロギーを抱かせるに至った要因を追求する.

本稿で観測したイデオロギー間の共起の頻度を可視化したグ

ラフを表 3 に示す. この図は bi, tri, quadri ユーザがイデオロギー A とイデオロギー B とをともに有していた場合に, その 2 つのイデオロギー間にエッジが張られたものであり, エッジの太さはその頻度を表す. また, 各ノードの位置 (レイアウト) はグラフ可視化ソフトウェア Gephi の ForceAtlas2 アルゴリズムにより計算され, 2 つのノード間の距離が短いことは, それらのイデオロギーの共起頻度が大きいことを示す. 更にこのアルゴリズムによって共起するイデオロギーのグループを見るのが可能になる. 具体的に言うと, この図から, 「野党」グループ (同図左) と 「与党・維新」グループ (同図右) とに大別されていることが分かる. さらに, それら 2 つのグループを直接的に結ぶエッジはほぼ存在しないことや, MT イデオロギーを介して間接的にこれら 2 つのグループが繋がっていることも分かる.

この図を踏まえて, 対立するイデオロギーを有するかどうかでユーザ群を分類し, 比較調査する. 具体的には, 「与党・維新」グループのイデオロギーと 「野党」グループのイデオロギーとをそれぞれ抱えている群を 「対立イデオロギー群」, 対していずれか一方のイデオロギーだけを抱えている群を 「単一イデオロギー群」として比較する.

6 トピックごとに異なるイデオロギーを持つユーザの分析

本節では, 対立するイデオロギーを抱えているユーザを分析し, 対立イデオロギーを抱かせるに至った要因を追求する. 前節で抽出したユーザのうち, 「与党」イデオロギーと 「野党」イデオロギーのいずれか少なくとも 1 つを共に保持していたユーザを抽出し, 各トピック内における投稿内容やリアクション先

の政治家, RT/MT/QT の内容などを確認した結果, 以下のよう
な要因がトピックによるイデオロギー変容の背景に存在する
と推測された。

共通の意見 これは, Twitter ユーザが, 全体的なイデオロギー
は異なるが特定のトピックにおいては共通の意見を有する政治
家の投稿に対してリアクションを行ったため, イデオロギーの
変容と認識されたケースである。この事例は特に表現の自由
トピックにおいて確認された。具体的には, 基本的なイデオロ
ギーは野党支持で与党反対派であるが, 「表現の規制」に対し
て反対意見を表明しており, 同じく表現規制に異を唱えている与
党議員に対して例外的に RT を行うユーザが多く存在した。

専門性 これは, 特定のトピックに高度の専門性を持つと考え
られる政治家の投稿に対してリアクションを行ったため, イデ
オロギーの変容と認識されたケースである。この事例は特に経
済トピックと外交トピックという, 他のトピックと比べて比較
的国内での意見が統一されているトピックで見られたことも注
目すべき点である。例えば, 基本的なイデオロギーは野党支持
で与党反対派であるが, 経済トピックに関心があり, 経済に詳
しい議員であれば与党議員でも例外的に RT するユーザが存在
した。

皮肉 これは, 支持ではなく皮肉の目的で対立するイデオロ
ギーを持つ議員の投稿を RT するケースである。この事例は多
くのトピックで見られたが特に改憲トピックにおいて高頻度に
確認されたことは特筆に値する。なお, このケースは実際には
イデオロギーを変容させたわけではなく, 単に賛同による RT
と批判を目的とした RT を区別していないがために検知された
ものであるが, 自身と異なるイデオロギーの情報に接触してい
るという点では重要と考えられるため記載する。

これらのケースのうち, 「共通の意見」の例から考えると, 仮
に自分とは対立するイデオロギーを有していたとしても, それ
が必ずしもすべての面での対立を意味するわけではないことが
伺える。また, 特に対立イデオロギーを有する人間であっても
同調できる面もあればイデオロギーの変容を喚起しうることが
推測される。同様に, 高度の専門性を有していれば, 必ずしも
イデオロギーの対立が情報交換を阻害するとは限らないことも
示唆されている。しかしながら, いずれにせよ, 今回はサンプ
ル数が少なく, 一般化するのは早計であると考えられるため,
今後の方針としてまず, サンプル数を増やしてより一般的な推
察を行いたい。

文 献

- [1] N. Gillani, A. Yuan, M. Saveski, S. Vosoughi, and D. Roy, “Me, my echo chamber, and i: Introspection on social media polarization,” in *Proceedings of the 2018 World Wide Web Conference*, WWW '18, (Republic and Canton of Geneva, CHE), p. 823–831, International World Wide Web Conferences Steering Committee, 2018.
- [2] A. Spitz, A. Abu-Akel, and R. West, *Interventions for Softening Can Lead to Hardening of Opinions: Evidence From a Randomized Controlled Trial*, p. 1098–1109. New York, NY, USA: Association for Computing Machinery, 2021.
- [3] J. Zeng, J. Li, Y. He, C. Gao, M. Lyu, and I. King, *What Changed Your Mind: The Roles of Dynamic Topics and Discourse in Argumentation Process*, p. 1502–1513. New

York, NY, USA: Association for Computing Machinery, 2020.

- [4] 大前拓巳, 豊田正史, 張翔, and 吉永直樹, “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,”
- [5] P. Barberá, J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau, “Tweeing from left to right: Is online political communication more than an echo chamber?,” *Psychological Science*, vol. 26, no. 10, pp. 1531–1542, 2015.
- [6] C. Monti, G. Manco, C. Aslay, and F. Bonchi, “Learning ideological embeddings from information cascades,” *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, Oct 2021.
- [7] X. Cheng, X. Yan, Y. Lan, and J. Guo, “Btm: Topic modeling over short texts,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 2928–2941, 2014.
- [8] T. Lin, W. Tian, Q. Mei, and H. Cheng, “The dual-sparse topic model: Mining focused topics and focused terms in short text,” in *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, (New York, NY, USA), p. 539–550, Association for Computing Machinery, 2014.
- [9] K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell, “Text classification from labeled and unlabeled documents using em,” *Machine learning*, vol. 39, no. 2, pp. 103–134, 2000.
- [10] J. Qiang, Z. Qian, Y. Li, Y. Yuan, and X. Wu, “Short text topic modeling techniques, applications, and performance: A survey,” *IEEE Transactions on Knowledge & Data Engineering*, pp. 1–1, may 5555.

謝 辞

本研究は JST CREST JPMJCR19A4, および, JSPS 科研費 JP21H03445 の支援を受けたものです。