

# SNS 上における非友好的リプライ行動の解明： 社会ネットワーク形成の観点から

代表研究者	山田 順子	立正大学 心理学部 助教
共同研究者	佐藤 浩輔	バンダイナムコ研究所 研究員
共同研究者	中分 遥	安田女子大学 心理学部 講師

## 1 研究の背景

現代社会において、SNS(ソーシャル・ネットワーク・サービス)は対人関係の形成や維持の重要なツールとして機能している。SNS の中には、Facebook や Mixi のようにユーザのオフライン上の対人関係を反映する閉じたコミュニティもあれば(Thomson & Ito, 2012)、Twitter や Instagram のように不特定他者に対して開かれたコミュニティもある。こうした様々な SNS の中で、本研究は特に Twitter に注目する。Twitter はオンライン・ソーシャル・ネットワーク研究において最も広く用いられるコミュニティの一つである(Aarts et al., 2012)。日本では世界的に見ても Twitter のアクティブユーザ数の多い国であることが知られている。日本における Twitter アクティブユーザ数は 2022 年 1 月時点で 58.95 万人であり、これは世界一位のアクティブユーザ数を持つアメリカ 76.90 万人に次いで 2 番目に多い数である(Dixson, 2022)。また日本国内における Twitter の利用率は 46.2%と高く、不特定他者に対して開かれたコミュニティとしては最も多く利用されている SNS の一つである(総務省, 2022)。このことから、本研究では日本における SNS 上の行動を分析するにあたって Twitter に着目した。

Twitter を始めとする SNS の発達に伴い、かつて対面で行われていたコミュニケーションは、SNS を介したオンライン・コミュニケーションへと切り替わりつつある(Liberman & Schroeder, 2020)。先行研究は、オフライン・コミュニケーションとオンライン・コミュニケーションとを比較し、それぞれのコミュニケーションにおける構造的違いとして以下の 4 つを挙げている(Liberman & Schroeder, 2020):1) 非言語的の手がかりの不足、2) 潜在的な匿名性の高さ、3) 新しい紐帯の形成と弱い紐帯の強化、そして 4) 情報の幅広い拡散である。こうした構造的な違いは、しばしば SNS 特有の新たな問題を生み出し、SNS ユーザの精神的健康に悪影響を与えることが指摘されている(e.g., Ojiha et al., 2021; Jeri-Yabar et al., 2019; Pantic, 2014)。

例えばオンライン・コミュニケーションでは、その性質上対面のコミュニケーションで利用される表情や声色といった非言語的の手がかりが欠落する(Dunlap & Lowenthal, 2009; Kiesler et al., 1984)。こうした非言語的の手がかりの不足は、情報の送り手側の考えや感情に対する情報の受け手側の推測精度の低下をもたらす(Hall & Schmitt Mast, 2007; Kruger et al., 2005)。このため、SNS などのオンライン・コミュニケーションにおいては、オフライン・コミュニケーションに比べて情報の送り手側の意図が正確に伝わりにくくなり、情報の送り手側の意図が誤って解釈されやすい。それゆえに、SNS 上でのコミュニケーションは齟齬が生じやすく、その結果として情報の送り手の本来の意図や文脈から外れた攻撃的な反応(非友好的リプライ行動)が引き起こされやすくなる。こうした SNS 上の本来の意図から外れた他者からの反応や返信は、投稿者の心理的ストレスとなることが指摘されている(植田ら, 2015; Walther & D'Addario., 2001)。

他にも、個々の SNS ユーザを超えた社会的問題として、ある対象に対して誹謗や中傷が集中し収まりがつかなくなる「炎上」(Aiken & Waller, 2000; Rajapaksha et al., 2019; Rosenberg, 2017; 山口, 2015)や、フェイクニュースなどの誤った情報の拡散(Del Vicario et al., 2016)、政治的態度や価値観の違いによって引き起こされる社会的分断(Adamic & Glance, 2005; 小林, 2012; 笹原・杜, 2019)などが知られる。

こうした社会的問題の中でも、SNS における「炎上」は、異なる文化圏ないしは価値観が SNS を介して接触することで生じうると考えられる。なぜなら、特定の文化圏・職業・分野といったカテゴリに属する個人にとっては当然の常識として共有されていることが、別のカテゴリに属する個人には共有されていないという齟齬がしばしば起こるからである。それゆえに、異なる文化的背景を持つ人同士が接触すると、しばしば外集団、すなわち同じ文化的背景を共有していない人々の行動は規範を逸脱した行動と解釈され、攻撃や非難といった非友好的リプライ行動が向けられやすくなる。これまで心理学では、文化的背景の違いが人々の行動・認知・価値観に影響を与えることを示しており(Tindale, 1995; Markus & Kitayama, 1991)、異なる文化的背景を持つ個人同士が相互作用することで文化的摩擦が生じることを示している(Bourhis et al., 1997; Sam & Berry, 2010)。

特に Twitter や Instagram といった不特定多数に開かれた SNS コミュニティでは、最低限のフィルタリングを行っているものの、ユーザが生成したコンテンツに対して他のユーザが反応するという性質からシステ

ム側がユーザに対して非友好的リプライ行動を取らないよう直接的に介入することは難しい。そのため、異なる文化的背景を持つユーザ同士の接触によってもたらされる問題については、まず SNS の投稿に対して人々がどのような振る舞いをするのか理解する必要がある。

SNS 上の投稿に対する振る舞いは、大きく 2 つのレベルに大別できる。第一に、他者が SNS に投稿した内容に対して人々はどのように振る舞うのかと、個人レベルの心理・行動の問題である。第二に、SNS 上の投稿に対する行動にはどのようなパターンがあるかといった、集団レベルの特性の問題である。例えばネットの炎上は、1)ある人が炎上に参加するかどうかを左右する心理的要因とはなにか、そして 2)炎上に参加している人々の行動にはどのようなパターンがあるのか、という 2 つの視点から理解できるだろう。SNS 上の投稿に対する人々の振る舞いを 2 つに大別したとき、人々が他者の投稿に対してどのような振る舞いをするかといった個人レベルの心理・行動の問題に関しては、質問紙などを用いた実験的検討が可能である。一方で、SNS 投稿に対する集団レベルの行動パターンの検討に関しては、積極的に SNS 上で投稿や投稿に対して反応するユーザの数が限られていることもあり、実験的検討が困難である。例えば SNS の炎上に関わるユーザは全体のごく一部であるという指摘があり (cf. 山口, 2015)、実際に炎上に関わっているユーザを集めて実験的な検討を行うことは現実的に困難である。そのため、集団レベルで人々の SNS 上での行動にどのようなパターンがあるかを検討するためには、現実の SNS 利用に関するデータの分析が必要となる。

将来的には、個人レベルと集団レベルの両方に着目した検討が必要であるが、本研究ではまずその端緒として、実際の SNS 上のツイート (Twitter に投稿される 140 以下の文章) について、ツイートの内容とマクロレベルでのユーザの反応との関係を検討する。具体的には、Twitter 上の特定のツイート群から共通する主題を抽出し、その主題に対する反応にどのようなパターンが見られるのかを検討する。

Twitter 上のある投稿に対してユーザができる直接的な反応には、自分自身のフォロワーに向けて投稿を拡散する「リツイート」、気に入った投稿に対しハートマークをつける「いいね」、投稿に対して直接返信を送る「リプライ」、そして投稿にコメントをつけた上で自分自身のフォロワーに向けてツイートを拡散する「引用」がある。なお、Twitter の初期設定では自分自身の投稿に対して他のユーザが何らかの反応をした場合に通知がなされる。この内、特にリプライと引用については、元のツイートを投稿したユーザに対して直接メッセージの内容が伝わることから、非友好的リプライ行動による衝突や対立が起こりやすい。

本研究では、Twitter 上の投稿内容と他のユーザの反応との関連を調べることで、リプライ行動の特徴を明らかにすることを目的とする。また、ツイートの投稿テキストの感情極性を抽出し、ツイートの感情極性が元ツイートに対する反応とどのように関係しているかを調べる。さらに、元ツイートに対するリプライ行動自体がどのような特性を持つかを分析することで、Twitter 上における反応の連鎖の一端について明らかにする。

## 2 方法

### 2-1 対象

トピックおよび反応を評価する対象として、日本語の報道系アカウント (全国紙である朝日新聞、読売新聞、産経新聞、毎日新聞、および NHK ニュース) のツイートをを用いた。一般的なアカウントの「炎上」や「バズった」 (他のユーザから多くの反応を集めた) ツイートは、アカウント自身の特性や他のユーザとのやりとりといった個別かつ経路依存的な経緯があると考えられることから、ツイートの内容と反応との一般的な関係性を調べるのに適切ではないと考えられる。そこで、比較的中立であり、かつ扱うトピックが幅広く安定的に一定数ユーザからの反応が期待できる報道系のアカウントを対象とした。

データの取得には Twitter API v2 を用い、2021 年 6 月 1 日から 2022 年 5 月 31 日までに投稿されたツイートの内容を対象にデータセットを構築した。データの取得は 2022 年 7 月に行った。取得したツイートは総計 227,483 件 (朝日新聞 45,533 件、読売新聞 40,377 件、産経新聞 29,913 件、毎日新聞 53,144 件、NHK ニュース 58,516 件) であった。

### 2-2 手続き

#### (1) データ前処理

ツイートを文書の単位とし、Python (3.7.9) を用いてデータの前処理を行った。文書はまずハッシュタグおよび URL を除外した上で形態素解析を行い、名詞 (普通名詞および固有名詞) を抽出した。形態素解析には Python のモジュールである spaCy (3.4.4) (Honnibal et al., 2017) および GiNZA (5.1.2) (松田ら, 2019) を用いた。トピックの分類には重複を排除した結果を用いた。名詞抽出後、異なり語数が 5 以上の文書を用いた。トピックの分離を容易にするため、文脈に関わらず出現する一般的な名詞、日付や時間に関する名詞をストップワードとして除外した。

## (2) トピックモデル

文書集合の背後に存在する潜在的なトピックを抽出するために LDA(潜在ディリクレ配分法)を行った。LDAにはPythonのモジュールである gensim(3.6.0) (Řehůřek & Sojka, 2010)を用いた。学習用データとしてアカウントごとに全文書の10%を抽出してモデルを学習させて分類器を作成し、残りのデータセットに適用するという方式を用いた。なお本研究の主眼は分類そのものではないため、学習に用いたデータも加えて後の分析に用いた。

文書集合に対してトピック数を5から200に変化させてグリッドサーチを行い、perplexityとcoherence(UMass)を指標として最適なトピック数を求めた。次いで、各トピックにおける代表的な単語を抽出するために、各トピックで上位の重みをもつ単語群、および分類後の各トピックに属する文書における上位頻出語を出力した。トピックの分離を妨げている単語があった場合はストップワードに追加して、再度トピック数の決定および代表的な単語の抽出を行った。

## (3) 極性分析

投稿の極性がユーザの振る舞いに与える影響を調べるために、辞書ベースの極性分析を行った。辞書には高村ら(2006)の単語感情極性対応表(Semantic Orientations of Words)を用いた。これは日本語55,125件について-1~1の感情極性を割り振ったものである。形態素分析を行って見出した単語について、高村らの単語感情極性対応表から極性を求め、それぞれの投稿ごとに極性を集計した。数詞、助詞、句読点、記号および分類不能の品詞については事前に除外した。

## (4) リプライ行動の分析

データの収集は2023年1月および2023年4月に行った。リプライについてはTwitter API v2のFull Archive Search APIを、引用についてはQuote Tweets APIを用いて、最終的に663,680件のリプライ、618,660件の引用ツイートを収集した(注1)。各投稿についても極性分析を行い、感情極性値を得た。

リプライ行動の特性が、リプライ元となるツイートの特性によってどの程度説明できるかを調べるために、一般化線形混合モデルを用いて分析を行った。具体的には元ツイートに対するリツイート、いいね、リプライ、引用、極性およびカテゴリを説明変数、元ツイートに対する反応ツイートの各変数を目的変数として、分析を行った。ただし、リツイート、いいね、リプライ、引用の間にはそれぞれ強い相関があることから、ツイートを単位とした主成分分析で得られた主成分得点を説明変数に投入した。また、当該の反応ツイートがリプライ行動であるかどうかをダミー変数として投入した。モデルは元ツイートのIDおよび元ツイートの投稿者IDをランダム切片として投入した混合モデルであった。計数データであるリツイート、いいね、リプライ、引用数のモデルには、リンク関数としてゼロ過剰負の二項分布を用いた。分析にはR(4.2.3)のglmmTMB(1.1.7)およびlme4(1.1.32)を用いた。分析にはリプライ・引用の双方が存在するツイート73,524件と、そのリプライ・引用計1,263,747件を用いた。

## 3 結果

### 3-1 トピックの分類

LDAの結果、80のトピックが抽出された。このうち、内容の理解に寄与しない単語群からなる解釈不能なトピック、および全く異なる複数の主題が混合しているトピックを除外し、最終的に54のトピックを分析に用いた。各トピックについて大まかな上位カテゴリ(「政治(Politics)」「経済(Economy)」「新型コロナ(COVID-19)」「国際情勢(International affairs)」「国内情勢(Domestic affairs)」「事件・事故(Incidents/Accidents)」「裁判(Trial)」「自然災害(Natural disasters)」「スポーツ(Sports)」「文化(Culture)」)を設定し、トピックをグループ化した。各カテゴリ、トピック、およびトピックの代表的な単語を表1に示す。

表 1 分類されたトピック、カテゴリと代表的な例

Category	Topic	n
Politics	assemblyperson	14873
	cabinet / prime minister	5925
	(opposition) party	2333
	cabinet	1794
	ruling party	1638
Economy	interest rates	3037
COVID-19	infection(1)	7309
	infection(2)	6558
	vaccine	4682
	infection(3)	2727
	pre-emergency measures / state of emergency declaration	2697
	medical system	2368
	infection(4)	2271
	Omicron variant	1935
International affairs	the situation of Ukraine: invasion	5018
	US-China relations / Taiwan	3301
	US-Japan relations	2203
	China	2157
	the situation of Ukraine: military	2106
	North Korea	2095
	Heads of state	2067
	South Korea	2067
	Afghanistan	1800
	China / sea	1737
	Russia / sanction	1704
	USA	1611
	Government	1017
Domestic affairs	Okinawa	2650
	Hokkaido	1837
Trial	district court	2849
	judicial decision	2766
Incidents/Accidents	murder / arrest	5867
	operation	4107
	death	3918
	referral to prosecutors	2679
	corpse	2607
	Osaka building fire	1981
Natural disasters	Atami landslide	2973
	earthquake / tsunami	2379
	typhoon / landslide disaster	2191
Sports	tournament: championship	4944
	olympic: Beijin	4154
	olympic: medal	4119
	toumament	2908
	National High School Baseball Invitational Toumament	2768
	olympic: athlete	2743
	tournament: final	2729
	Japanese High School Baseball Championship	1815
	tournament: FIFA World Cup	1562
Culture	Sota Fujii(professional shogi player)	2944
	obituary	2817
	Go / Shogi(Japanese chess)	2457
	director	2358
	award / prize	1561

$n$ はトピックに分類された文書数。いずれかのカテゴリへの所属確率が0.5以上のものを抽出後、所属確率のもっとも高いトピックを分類結果とした。

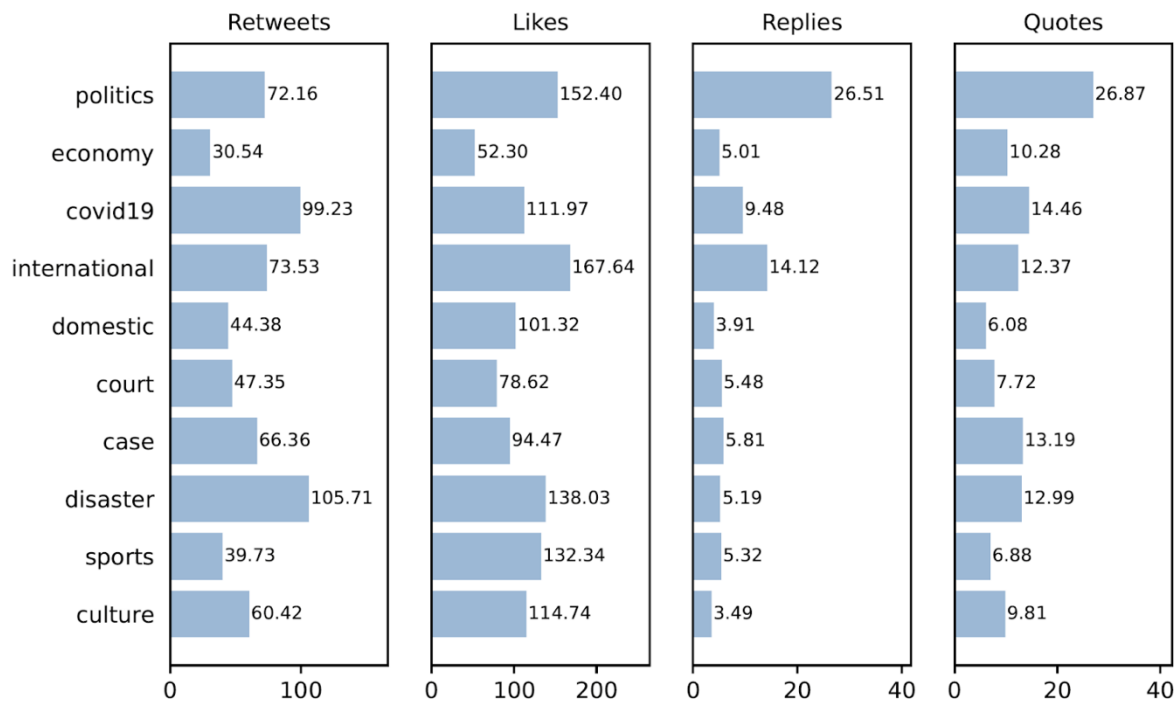
2019年に発生した新型コロナウイルス感染症(COVID-19)は、データ収集時期の2022年段階でも依然として世界的な流行が続いており、新型コロナウイルス感染症関連のトピックが複数抽出された。感染については複数のトピックが抽出されたが、頻出語の観点からは明瞭に区別できなかったため、便宜的に「感染1」～「感染4」と命名した。また2021年に東京オリンピック、2022年に北京オリンピックが開催されたことから、オリンピックに関するトピックが複数抽出された。国際事情に関しては、2022年2月にロシアによるウクライナ侵攻が始まったことから、ウクライナとロシアに関わるトピックが複数抽出された。その他、2021年7月に熱海市で起きた土石流災害、2021年12月に大阪市で起きたビル放火事件など、日本国内で大きく報道された災害や事件が個別のトピックとして抽出された。

### 3-2 カテゴリ、トピックとユーザの反応

カテゴリへの所属確率が0.5以上のツイートを抽出し、元ツイートに対するリツイート、いいね、リプライ、引用といった他ユーザの反応の平均値を算出した(表2)(注2)。その結果、カテゴリによって反応のパターンが異なることが明らかとなった。

政治カテゴリにおいては、いずれの反応も高い水準であり、特に引用とリプライが他のカテゴリに比べて顕著に多かった。新型コロナウイルス感染症や国際情勢といったカテゴリにおいては、リツイートやいいねは高い水準であるものの、リプライや引用は中程度の水準であった。また自然災害や事件・事故といったカテゴリにおいては、リツイートやいいねは高い水準であるが、リプライは比較的低水準であり、また引用は中程度であった。スポーツや文化カテゴリにおいては、いいねは高水準であるが、リツイートは中程度、リプライ・引用はいずれも低水準であった。

表2 カテゴリとユーザの反応(値は平均値を示す)



カテゴリレベルの反応の平均値を用いて主成分分析を行った(図1)。分析の結果、第一主成分は全体的な反応の量を示していると解釈できる(PC1:元ツイートの拡散度)。第二主成分については、リプライや引用といった発言を伴う反応なのか、あるいはリツイートやいいねといった発言を伴わない反応なのかを示していると解釈できる(PC2:元ツイートのリプライや引用のしやすさ)。元ツイートの拡散度という観点(PC1)からは、政治や新型コロナウイルス感染症、国際情勢、自然災害といったカテゴリで比較的反応の量が多く、経済や裁判、国内情勢といったカテゴリでは比較的反応の量が少なかった。また元ツイートのリプライや引用のしやすさ(PC2)という観点からは、政治や経済といったカテゴリでは発言を伴う反応(i.e.,リプライや引

用)が優勢である一方、自然災害や文化・スポーツ、国内情勢では発言を伴わない反応(i. e., リツイートやいいね)が優勢であった。全体的に見ると、政治カテゴリは反応の量が多くかつ発言を伴うものが多く、自然災害は反応の量が多いが発言を伴わないものが多かった。また文化・スポーツ・国内情勢は反応の量が少なく発言も伴わないものが多く、経済は反応の量は少ないが発言を伴うものが多いといった特徴をそれぞれ持っていた。

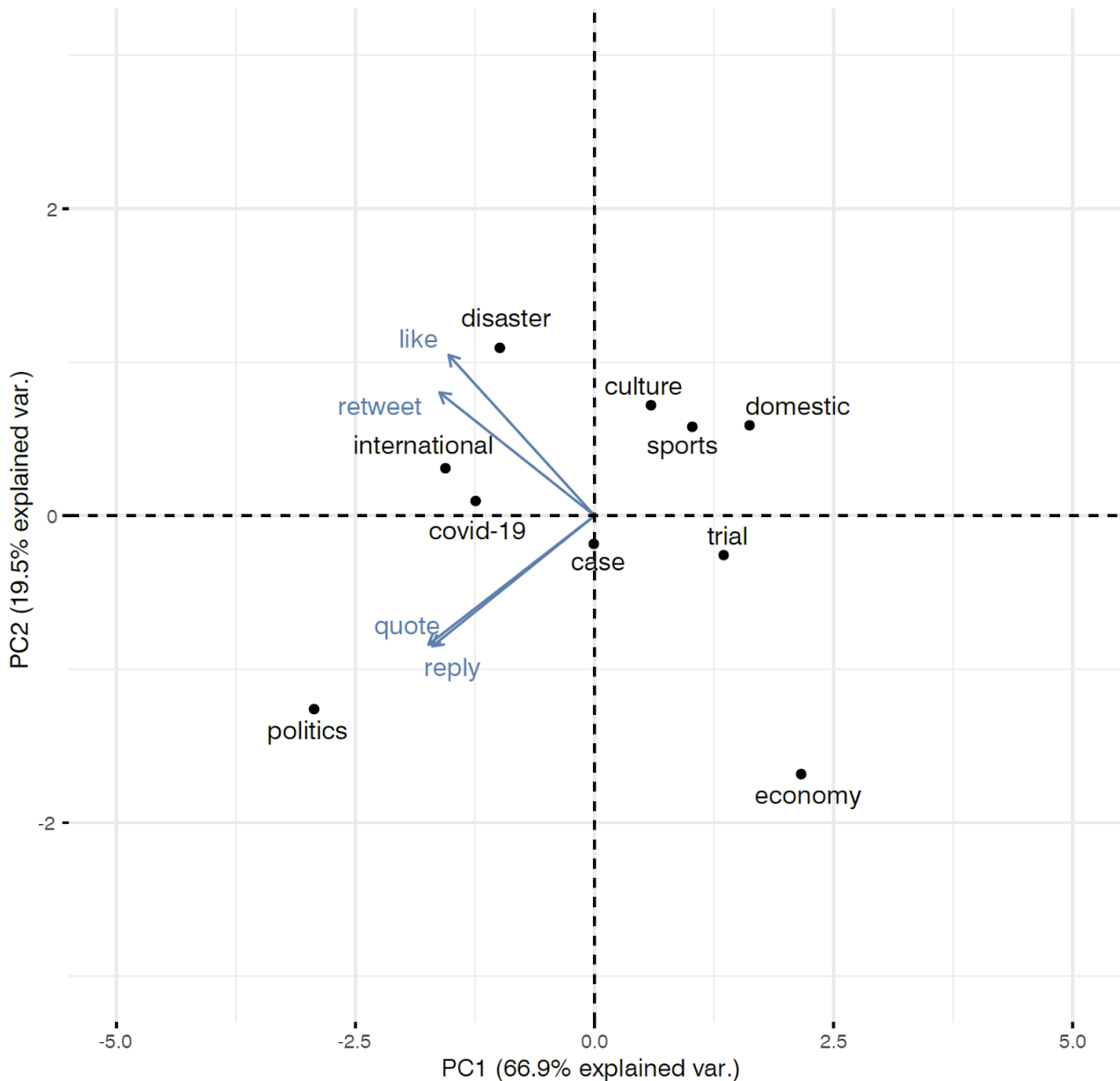


図1 カテゴリと反応のパターン(対数化した平均値を用いた主成分分析の結果)

トピックと反応のパターンについて調べるため、同様の分析をトピックについても行った(図2)。その結果、カテゴリの場合と同様、第一主成分に元ツイートの拡散度、第二主成分に元ツイートのリプライや引用のしやすさと解釈できる結果が得られた。一方、同じカテゴリに分類したトピックでも布置の異同がみられ、新型コロナウイルス感染症に関わる話題であっても、感染に関わるトピックではリツイートやいいねといった発言を伴わない反応が多かった。一方で、ワクチンやオミクロン株といったトピックではリプライや引用といった発言を伴う反応が多かった。また国際情勢に関わる話題でも、韓国や北朝鮮といったトピックでは

発言を伴う反応が多かった一方、台湾や中国、ウクライナ情勢に関わるトピックでは発言を伴わない反応が多かった。

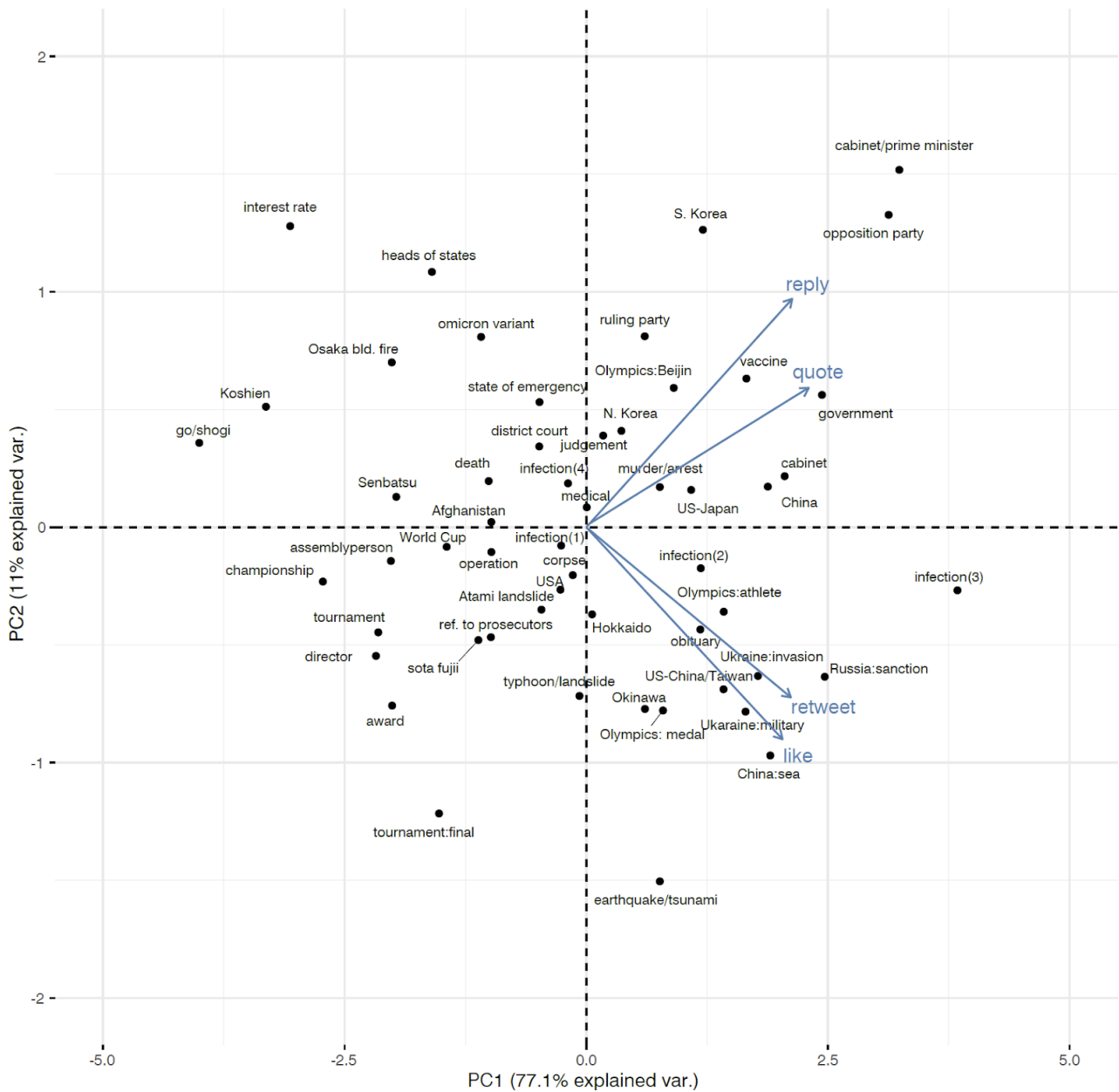


図2 トピックと反応のパターン(対数化した平均値を用いた主成分分析の結果)

### 3-3. カテゴリと感情極性

カテゴリと投稿の極性を調べるために、カテゴリごとの極性値を箱ひげ図で示した(図3)。極性値は正の値と負の値のいずれもとる可能性があるが、全体としてほとんどの投稿がゼロ以下に分布しており、極性値の分布は負の側に裾の長い歪んだ分布を示した。中央値で見ると、事件・事故に関するツイートがもっともネガティブであった。一方、新型コロナウイルス感染症や国内情勢に関するツイートは相対的にポジティブであることが示された。

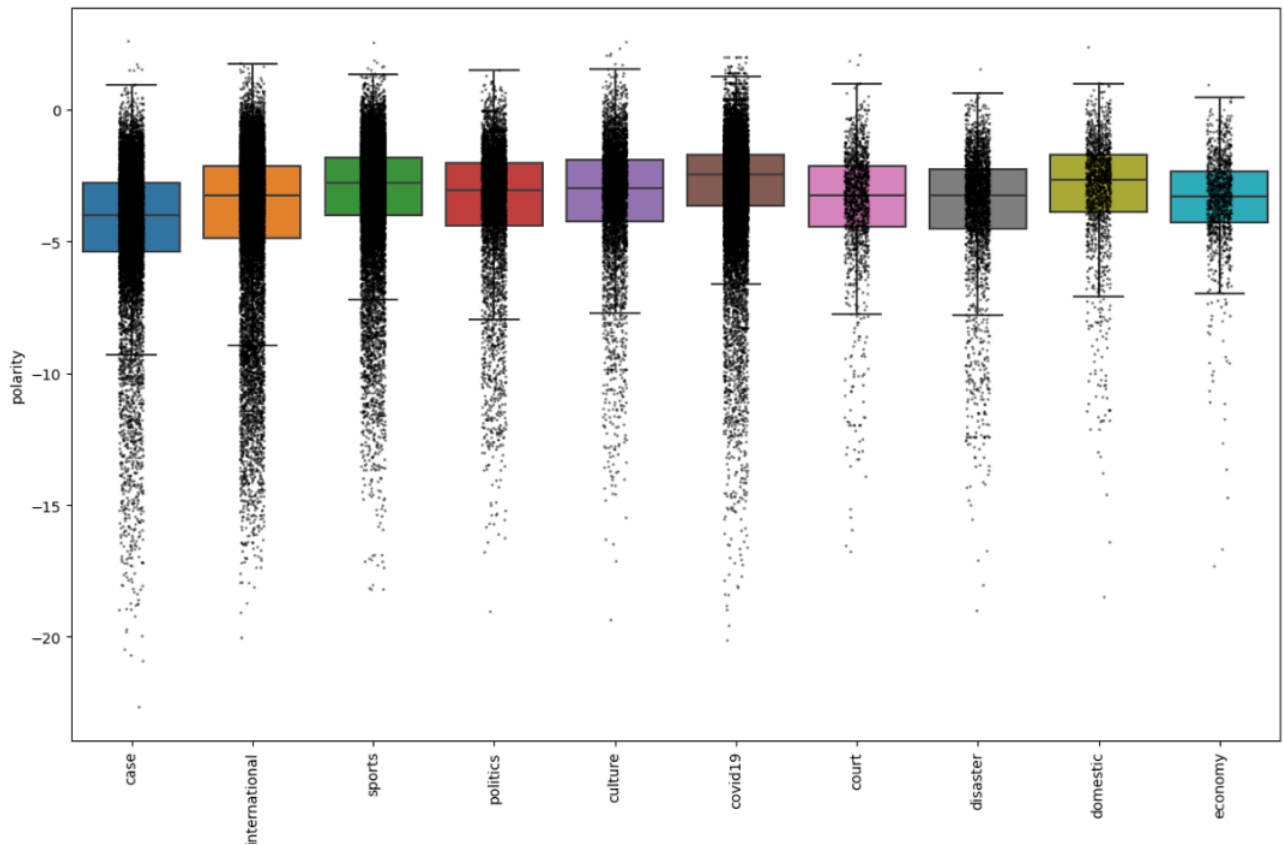


図3 ツイートのカテゴリごとの感情極性

### 3-4 反応ツイートの特性に対する元ツイートの特性の影響

反応ツイートの特性が元ツイートの特性によってどの程度説明できるかを調べるために、一般化線形混合モデル(General Linear Mixed Model: GLMM)を用いて分析を行った(表3)。分析の結果、リツイート、いいね、リプライ、引用数といった反応それぞれについて、元ツイートの拡散度(PC1)が有意な強い正の影響を持つことが示された。つまり、元ツイートが人気で多くのユーザに拡散されるほど、それに対する反応ツイートもまた多くの反応を得る傾向が示された。リツイート、いいね、引用については、元ツイートに対するリプライや引用のしやすさ(PC2)も有意な影響を持っていた。つまり、元ツイートがリプライや引用といった発言を伴う反応を引き起こしやすいほど、元ツイートに対する反応ツイート自体もまた、二次的な反応(元ツイートへの反応に対するさらなるリプライや返信)を引き起こしやすいことが示された。極性については、リツイート、いいね、引用で有意な負の影響があった。つまり、元ツイートがネガティブであるほど、反応の量も多くなることが示された。またダミー変数であるリプライかどうかも、すべての変数に対して強い負の影響を与えていた。つまり、反応がリプライであるときに比べて引用であるときの方がより多くの反応を得られることが示された。カテゴリについても元ツイートの有意な影響が示され、カテゴリによって各変数の反応の仕方が異なっていた。極性についてはおおむねリツイートの反応と同様の傾向であったが、元ツイートのリプライや引用のしやすさ(PC2)の影響が負であった。つまり、元ツイートが反応を引き起こしやすいほど、元ツイートに対する反応ツイートの極性はネガティブになりやすかった。また元ツイートの極性の正の影響がみられ、元ツイートがポジティブであれば反応ツイートもポジティブに、ネガティブであればネガティブになりやすいことが示された。



表 3 元ツイートの特性が反応ツイートの特性に与える影響(一般化線形混合モデルの分析結果)

predictor	Retweets		Likes		Replies		Quotes		Polarity	
	$\beta$	95% CI	$\beta$	95% CI	$\beta$	95% CI	$\beta$	95% CI	$\beta$	95% CI
PC1	1.170 ***	1.090 - 1.240	1.500 ***	1.430 - 1.560	0.638 ***	0.590 - 0.687	0.528 ***	0.463 - 0.593	0.388 ***	0.314 - 0.463
PC2	0.099 **	0.028 - 0.169	0.063 *	0.004 - 0.123	0.026	-0.021 - 0.073	0.088 **	0.025 - 0.150	-0.334 ***	-0.407 - -0.261
polarity	-0.039 **	-0.067 - -0.012	-0.068 ***	-0.088 - -0.048	-0.016	-0.035 - 0.004	-0.046 *	-0.082 - -0.009	0.287 ***	0.263 - 0.311
is_reply	-1.280 ***	-1.300 - -1.260	-0.501 ***	-0.511 - -0.490	-0.571 ***	-0.584 - -0.557	-1.260 ***	-1.290 - -1.220	-0.291 ***	-0.305 - -0.277
category										
case	—	—	—	—	—	—	—	—	—	—
court	0.390 ***	0.233 - 0.546	0.123 *	0.016 - 0.229	0.042	-0.070 - 0.154	0.324 **	0.108 - 0.540	-0.330 ***	-0.460 - -0.201
covid19	0.497 ***	0.431 - 0.563	0.017	-0.028 - 0.063	0.068 **	0.021 - 0.114	0.300 ***	0.209 - 0.391	-0.808 ***	-0.863 - -0.753
culture	-0.098	-0.208 - 0.013	0.088 *	0.015 - 0.162	-0.134 ***	-0.213 - -0.055	-0.231 **	-0.387 - -0.076	0.494 ***	0.404 - 0.584
disaster	0.315 ***	0.197 - 0.434	0.153 ***	0.072 - 0.234	0.022	-0.063 - 0.106	0.013	-0.154 - 0.180	-0.141 **	-0.240 - -0.041
domestic	0.083	-0.088 - 0.254	0.083	-0.030 - 0.197	-0.101	-0.224 - 0.023	0.007	-0.240 - 0.255	-0.039	-0.177 - 0.098
economy	0.033	-0.194 - 0.260	-0.065	-0.217 - 0.087	-0.032	-0.192 - 0.128	-0.069	-0.386 - 0.248	0.186 *	0.003 - 0.369
international	0.224 ***	0.158 - 0.290	-0.097 ***	-0.142 - -0.052	-0.036	-0.083 - 0.010	0.079	-0.011 - 0.170	-0.304 ***	-0.359 - -0.250
politics	0.590 ***	0.514 - 0.666	0.260 ***	0.206 - 0.313	0.021	-0.033 - 0.075	0.370 ***	0.270 - 0.471	-0.232 ***	-0.297 - -0.166
sports	-0.158 ***	-0.233 - -0.083	-0.005	-0.055 - 0.045	-0.245 ***	-0.299 - -0.192	-0.234 ***	-0.340 - -0.129	0.708 ***	0.647 - 0.769
conditional R <sup>2</sup>	0.300		0.357		0.191		0.191		0.119	
marginal R <sup>2</sup>	0.191		0.263		0.103		0.101		0.031	

\*p<0.05; \*\*p<0.01; \*\*\*p<0.001

CI = Confidence Interval

Zero-inflated negative binomial was used as link function for retweets, likes, replies, and quotes.

Pseudo-R2 were calculated for non-linear model

カテゴリによって各変数の影響力がどのように異なるのかを調べるため、GLMM を行った。交互作用項を投入したところ強い多重共線性の問題が生じたため(VIF > 10<sup>4</sup>)、カテゴリごとに分けて分析を行った。先ほどのモデルと同様、元ツイートの ID および元ツイートの投稿者 ID をランダム切片として投入した。

分析の結果、元ツイートの拡散度(PC1)はほぼどのカテゴリでも返信・引用ツイートへの反応を増加させていた(表 4)。一方で、元ツイートのリプライや引用のしやすさ(PC2)は、事件や国内情勢においては返信・引用ツイートへの反応に対して正の影響を与えていたが、国際情勢においては負の影響を与えていた。つまり、事件や国内情勢においては、元のツイートが返信や引用を引き付けているほど、その返信や引用に対してさらなる二次的な反応を引き付けやすかったのに対し、国際情勢ではむしろ元のツイートが返信や引用を引き付けているほど、その返信や引用に対する二次的な反応は抑制されていた。また、文化や災害といったカテゴリでは、元ツイートのリプライや返信のしやすさ(PC2)は有意な影響を持たず、二次的な反応を促進も抑制もしないことが示された。

また、返信・引用ツイートの感情極性は元ツイートの感情極性と正の相関を示したが、元ツイートの返信や引用のしやすさ(PC2)が高いほど多くのカテゴリでより反応はネガティブになる傾向があった。また返信であるほうが引用であるよりも反応が少なく、かつネガティブな内容になりがちであった。

表 4 元ツイートの特性が反応ツイートの特性に与える影響(一般化線形混合モデルの分析結果)

objective variable	predictor	case	court	covid19	culture	disaster	domestic	economy	international	politics	sports
		(n=144837)	(n=15956)	(n=304942)	(n=45854)	(n=32501)	(n=10733)	(n=13283)	(n=334017)	(n=216684)	(n=144940)
retweets	PC1	1.350 ***	1.150 ***	0.849 ***	0.764 ***	0.657 **	1.400 ***	1.550	1.180 ***	0.996 ***	1.500 ***
	PC2	0.449 ***	0.116	0.092	-0.075	-0.051	0.696 **	-0.458	-0.148 *	0.073	0.349 **
	polarity	-0.095 *	-0.252 **	0.007	-0.003	0.052	-0.040	-0.202	-0.032	-0.002	-0.024
	is_reply	-0.855 ***	-1.420 ***	-1.450 ***	-1.120 ***	-0.733 ***	-1.450 ***	-0.667 ***	-1.320 ***	-1.370 ***	-1.260 ***
likes	PC1	1.630 ***	1.270 ***	1.180 ***	1.170	1.020 ***	1.430	2.560	1.510 ***	1.250 ***	1.740 ***
	PC2	0.417 ***	0.196	0.106 *	-0.052	0.010	0.626	-1.050	-0.126 *	0.112	-0.099
	polarity	-0.044	-0.245 ***	-0.111 ***	-0.031	0.005	-0.032	-0.271	-0.029	0.001	-0.041
	is_reply	-0.038 **	-0.575 ***	-0.730 ***	-0.534	-0.292 ***	-0.753	-0.088	-0.487 ***	-0.513 ***	-0.530 ***
replies	PC1	0.716 ***	0.677 ***	0.556 ***	0.456 ***	0.550 ***	0.695 ***	0.911	0.654 ***	0.613 ***	0.775
	PC2	0.232 ***	0.126	-0.010	0.075	0.083	0.375 **	-0.364	-0.157 ***	0.011	0.108
	polarity	-0.034	-0.145 *	0.022	-0.026	-0.012	0.060	-0.172 *	-0.022	0.035	-0.013
	is_reply	-0.352 ***	-0.603 ***	-0.383 ***	-0.616 ***	-0.394 ***	-0.594 ***	-0.427 ***	-0.678 ***	-0.934 ***	-0.421
quotes	PC1	0.594 ***	0.842 ***	0.563 ***	0.225	0.369	0.714 ***	1.720 *	0.548 ***	0.565 ***	0.567 ***
	PC2	0.341 ***	-0.179	0.051	0.020	0.215	0.397 **	-1.090	-0.080	-0.092	0.242 *
	polarity	-0.128 *	-0.088	-0.028	0.013	0.014	-0.021	0.045	0.000	-0.010	0.095 *
	is_reply	-0.680 ***	-1.210 ***	-0.999 ***	-0.867 ***	-0.782 ***	-1.030 ***	-0.641 ***	-1.640 ***	-1.650 ***	-1.220 ***
polarity	PC1	0.016	0.002	0.472 ***	0.209	0.199	-0.377	1.590 *	0.256 ***	0.504 ***	-0.107
	PC2	-0.836 ***	-0.401 *	-0.411 ***	-0.270 *	-0.566 **	-0.653 **	-1.480 *	0.079	-0.104	-0.921 ***
	polarity	0.432 ***	0.451 ***	0.221 ***	0.313 ***	0.323 ***	0.222 **	0.263 ***	0.215 ***	0.214 ***	0.193 ***
	is_reply	-0.266 ***	-0.187 **	-0.719 ***	-0.333 ***	-0.388 ***	-0.215 **	-0.326 ***	-0.090 ***	-0.015	-0.240 ***

\*p<0.05; \*\*p<0.01; \*\*\*p<0.001

Zero-inflated negative binomial was used as link function for retweets, likes, replies, and quotes.

#### 4 考察

本研究では、Twitter 上におけるリプライ行動、すなわち他ユーザの投稿に対するリプライや引用といったリプライ行動について、ユーザが投稿するツイートの内容やそれに対する反応パターンの特徴を明らかにすることを目的とし、実際に Twitter に投稿されたツイートからトピックを抽出し、各トピックとその反応（リツイート・いいね・リプライ・引用）の関係について検討した。

分析の結果、Twitter に投稿されたツイートのトピックによって引き起こされる反応パターンが異なることが明らかとなり、全体的な反応の量および、反応が発言を伴うかどうかによってツイートのトピックが特徴づけられることが示された。特に反応量が多いトピックの中でも、政治(内閣・首相、政党・野党)や新型コロナウイルス感染症対策(ワクチン、まん延防止・緊急事態宣言)など、様々な意見や立場を含むトピックについては発言を伴う反応が多い傾向があった。その一方で、自然災害(地震・津波、台風・土砂災害)や新型コロナウイルス感染症(感染)、国際情勢(ウクライナとロシア、中国と台湾)などといった危機や脅威に関するトピックについては発言を伴わない反応が多い傾向が示された。

これらの傾向は、政治のように個々人で考えや立場が異なるようなトピックはユーザ間の議論を生じやすいという「論争的な価値」、自然災害のような危機や脅威に関するトピックは関連する情報を広く周知するという「情報的な価値」という、投稿内容の質的な違いを示していると解釈できる。ユーザの視点から考えると、情報的な価値を持つツイートは、周辺他者に情報を伝達したり自身で情報を活用することを目的としてリツイートやいいねなど比較的受け身的な行動をとる一方、論争的な価値を持つツイートについては、議論に参加、つまり「口を挟む」ためにリプライや引用といった能動的な行動をとり積極的に自らの考えや立場を表明するという可能性が考えられる。このことは、異なる価値を持つ投稿の拡散にそれぞれ異なるメカニズムが働いている可能性を示唆する。すなわち、情報的な価値を持つツイートはリツイートを通して直線的に広がっていくのに対し、論争的な価値を持つツイートは元のツイートに対して様々なユーザがリプライや引用を行うことでさらに論争的な価値が高まり、他ユーザからの発言を伴う反応や、そうした反応に対する二次的な反応を引き起こしやすくなっていくというフィードバックがあると考えられる。

返信・引用ツイートの分析では、元のツイートのトピックや極性などの特性が返信・引用ツイートへの反応に影響を与えていることが示された。元のツイートの拡散度が高く他ユーザから多くの反応を集めているほど、また元のツイートが他ユーザからのリプライや引用といった発言を伴う反応を引き起こしやすい内容であるほど、加えて元ツイートの内容がネガティブであるほど他ユーザからの反応が増えることが示された。また元のツイートが多く反応を集めたり、元ツイートの内容がポジティブであるほど、元ツイートに対する返信や引用もまたポジティブな内容になる傾向があった。一方で、元のツイートがリプライや引用といった発言を伴う反応を引き起こしやすい内容である場合には、反応ツイートの内容はネガティブになる傾向が合った。また、元のツイートのトピック(カテゴリ)によって反応に対する影響力のパターンが異なることが示された。

このことは特定の性質を持つツイートが自己増幅的なプロセスによって拡散していくことを示唆している。すなわち、元のツイートが反応を集めやすく、かつネガティブでリプライや引用といった発言を伴う反応を引き起こしやすい内容である場合、その返信や引用もまたネガティブな内容になりやすく、かつ他ユーザからの二次的な反応を得やすいということである。こうしたプロセスが、Twitter における非友好的リプライ行動や、それに基づく炎上、ないしはユーザの断絶に寄与している可能性は無視できない。今後このような連鎖の構造についてより詳細に明らかにしていく必要があるだろう。

また、SNS におけるリプライや引用といったリプライ行動を「話題への参加」という観点でとらえることは有益であるだろう。なぜなら、SNS 上の行動という文脈特異的な行動から、より人間の一般的な心理・行動の問題へと還元することで、既存の社会心理学的研究の知見を援用しうるためである。伝統的な心理学の観点からは、SNS 上における衝突や分断、炎上の背後にあるメカニズムとして、しばしばユーザの悪意や敵意といった個人の内的特性が想定される。しかし、本研究の結果は、元ツイートに対するネガティブな反応といった非友好的リプライ行動が、悪意や敵意に基づいたものではなく、個々のユーザにとっては単なる話題への参加程度の動機に過ぎない可能性を示唆している。もしそうであるとすれば、議論の場をファシリテーターが整えていくように、SNS 上のコミュニケーションの構造を適切に設計することで、より平和的・友好的な事態の経過を期待できるかもしれない。

異なる文化的背景を持つ人々同士の接触は、非友好的リプライ行動や、衝突・対立・分断といったネガティブな結果をもたらす可能性を伴う。しかし、それと同時に、新たなネットワークの広がりや新規他者との交流の場を作り出すといったポジティブな可能性も期待される。近年、SNS に関してはネガティブな側面が取り沙汰されがちであるが、時空間を超えたネットワークの拡張は SNS が持つ元来的な価値のひとつである。こうした可能性を改めて見出したことは、本研究の意義の一つであるだろう。

一方で、本研究にはいくつかの限界がある。まず、トピック抽出の問題である。Twitter 上で投稿されるツイートには 140 字の字数制限があり(注 3)、それゆえに各文書に含まれる名詞数は少なくなる。このため、本研究の分析でもトピックが混交して抽出されたり、解釈が困難なトピックが現れたりといったことが見られた。そのため、文書集合の特徴を平均的には捉えているものの、細部に関しては適切でない分類が行われている可能性がある。また、トピックとして抽出されていない日常的な話題などの重要な主題が洩れているという問題もある。加えて、感情極性についても辞書ベースで集計した単純な合計であり、どれくらい個々のツイートが持つ感情極性を捉えきれているかは不明瞭である。これらについては、機械作業ではなく人間が行ったコーディングの結果を教師データとして分類器を作成しツイートの種類を分類するといった手法と組み合わせることで、結果の妥当性を担保していく必要があるだろう。また、本研究では報道系アカウントのツイートとその反応を対象に分析を行った。しかし、報道系アカウントのツイートは Twitter 上で観察される多様なツイートの部分集合に過ぎず、本研究で扱わなかった種類のツイートにも今回の分析で得られた知見が適用可能であるかどうかについては今後議論の余地がある。この点については、より広い範囲のツイートを対象に同様の分析を行うことで、より知見の頑健性を検証することが求められるだろう。

### 【参考文献】

- Aarts, O., van Maanen, P.-P., Ouboter, T., & Schraagen, J. M. (2012). Online social behavior on Twitter: A literature review 12th International Conference on Data Mining Workshops, (Vol. 2012) (pp. 739–746). IEEE Publications. <https://doi.org/10.1109/ICDMW.2012.139>.
- Adamic, L. A., & Glance, N. (2005). The political blogosphere and 2004 U.S. election divided them into blogs. In *Proceedings of the 3rd International Workshop on Link Discovery* (pp. 36–43). <https://doi.org/10.1145/1134271.1134277>.
- Aiken, M., & Waller, B. (2000). Flaming among first-time group support system users. *Information and Management*, 37, 95–100. [https://doi.org/10.1016/S0378-7206\(99\)00036-1](https://doi.org/10.1016/S0378-7206(99)00036-1).
- Bourhis, R. Y., Moïse, L. C., Perreault, S., & Sénécal, S. (1997). Towards an interactive acculturation model: A sociopsychological approach. *International Journal of Psychology*, 32, 369–386. <https://doi.org/10.1080/002075997400629>.
- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). Spread of misinformation online. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 554–559. <https://doi.org/10.1073/pnas.1517441113>.
- Dixon, A. (2022). Leading countries based on number of Twitter users as of January 2022(in millions). Statista. Retrieved from <https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>. Accessed 24/6/2023.
- Dunlap, J. C., & Lowenthal, P. R. (2009). Tweeting the night away: using Twitter to Enhance Social Presence. *Journal of Information Systems Education*, 20, 129–136.
- Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion*, 7, 438–446. <https://doi.org/10.1037/1528-3542.7.2.438>.
- Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: industrial-strength natural language processing in Python.
- Jane, E. A. (2015). Flaming? What flaming? The pitfalls and potentials of researching online hostility. *Ethics and Information Technology*, 17, 65–87. <https://doi.org/10.1007/s10676-015-9362-0>.
- Jeri-Yabar, A., Sanchez-Carbonel, A., Tito, K., Ramirez-delCastillo, J., Torres-Alcantara, A., Denegri, D., & Carreazo, Y. (2019). Association between social media use (Twitter, Instagram, Facebook) and depressive symptoms: are Twitter users at higher risk? *International Journal of Social Psychiatry*, 65, 14–19. <https://doi.org/10.1177/0020764018814270>.
- Kayany, J. M. (1998). Contexts of uninhibited online behavior: flaming social NewsGroups on Usenet. *Journal of the American Society for Information Science*, 49, 1135–1141. [https://doi.org/10.1002/\(SICI\)1097-4571\(1998\)49:12<1135::AID-ASI8>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1097-4571(1998)49:12<1135::AID-ASI8>3.0.CO;2-W).
- Kelly, L., & Miller-Ott, A. E. (2018). Perceived miscommunication in friends' and romantic partners' texted conversations. *Southern Communication Journal*, 83, 267–280. <https://doi.org/10.1080/1041794X.2018.1488271>.
- Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Socio-psychological aspects of computer-mediated communication. *American Psychologist*, 39, 1123–1134. <https://doi.org/10.1037/0003-066X.39.10.1123>.
- 小林哲郎. (2012). ソーシャルメディアと分断化する社会的リアリティ. *人工知能*, 27(1), 51-58.

- Kruger, J., Epley, N., Parker, J., & Ng, Z. W. (2005). Egocentrism through e-mail: can we communicate and think? *Journal of Personality and Social Psychology*, 89, 925–936. <https://doi.org/10.1037/0022-3514.89.6.925>.
- Lieberman, A., & Schroeder, J. (2020). Two social lives: how do differences between online and offline interactions influence social outcomes? *Current Opinion in Psychology*, 31, 16–21. <https://doi.org/10.1016/j.copsyc.2019.06.022>.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychological Review*, 98, 224–253. <https://doi.org/10.1037/0033-295X.98.2.224>.
- 松田寛・大村舞・浅原正幸. (2019). 短単位品詞の用法曖昧性解決と依存関係ラベリングの同時学習. 言語処理学会第 25 回年次大会発表論文集, 201-204
- Ojha, K., Soohinda, G., Sampath, H., & Dutta, S. (2021). Social networking sites and their relationships with social comparison and psychological well-being among Medical University students. *Indian Journal of Psychiatry*, 63, 593–596. [https://doi.org/10.4103/indianjpsychiatry.indianjpsychiatry\\_1344\\_20](https://doi.org/10.4103/indianjpsychiatry.indianjpsychiatry_1344_20).
- Pantic, I. (2014). Online social networking and mental health. *Cyberpsychology, Behavior and Social Networking*, 17, 652–657. <https://doi.org/10.1089/cyber.2014.0070>.
- Rajapaksha, P., Farahbakhsh, R., Crespi, N., & Defude, B. (2019). Uncovering Flaming Events in News Media in social media 38th International Performance Computing and Communications Conference (IPCCC), 1–8, (Vol. 2019). IEEE Publications. <https://doi.org/10.1109/IPCCC47392.2019.8958759>.
- Řehůřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50).
- Rosenberg, A. (2017). Nasty discussion: an overview of flaming and incivility in news comment sections. *IAMCR 2017 Conference*, Cartagena, Colombia.
- Sam, D. L., & Berry, J. W. (2010). Acculturation: when individuals and groups from different cultural backgrounds meet. *Perspectives on Psychological Science*, 5, 472–481. <https://doi.org/10.1177/1745691610373075>.
- 笹原和俊・杜宝発. (2019). ソーシャルメディアにおける道徳的分断: LGBT ツイートの事例. *社会情報学*, 8(2), 65-77.
- 総務省情報通信政策研究所. (2022). 令和 3 年度情報通信メディアの利用時間と情報行動に関する調査報告書.
- 高村大地・乾孝司・奥村学. (2006). スピンモデルによる単語の感情極性抽出. *情報処理学会論文誌ジャーナル*, 47(2), 627-637.
- Thomson, R., & Ito, N. (2012). The effect of relational mobility on SNS user behavior: A study of Japanese dual-users of Mixi and Facebook. *Journal of International Media, Communication, and Tourism Studies*, 14, 3–22.
- Triandis, H. C. (1995). *Individualism and collectivism*. Westview Press.
- 植田智明・折原良平・清雄一・田原康之・大須賀昭彦. (2015). ユーザーの期待する反応に応じたツイート分類. 第 29 回人工知能学会全国大会. 於: 公立はこだて未来大学.
- Walther, J. B., & D'Addario, K. P. (2001). The impacts of emoticons on message interpretation in computer-mediated communication. *Social Science Computer Review*, 19, 324–347. <https://doi.org/10.1177/089443930101900307>.
- 山口真一. (2015). ネット炎上の実態と政策的対応の考察—実証分析から見る社会的影響と名誉毀損罪・制限的本人確認制度・インターネットリテラシー教育の在り方—. *情報通信政策レビュー*, 11, 52-74.

(注1) 元ツイートのリプライ数、引用数の合計はそれぞれ 695,348 件、916,137 件であり、実際に取得できた数の間に差があった。これは鍵アカウントの投稿が収集出来なかったことや、データの収集時期が影響した可能性がある。元ツイートに対する反応データを収集していた 2023 年 1 月～4 月の間には Twitter 上でアカウントの大量凍結が行われ、また Twitter からユーザの流出が起こっており、こうしたことがデータの収集に影響した可能性が考えられる。

(注2) 無作為につぶやきを選出した際の期待値という観点から、平均値を代表値として用いた。中央値は外れ値に対して頑健であるという性質を持つが、同時に極端な値の情報が失われてしまうため、Twitter のツイートのように反応量が極端に多いツイートがある程度の代表性を持つケースでは必ずしも適切ではないと考えられる。

(注3) 2021年6月から有料サブスクリプションである「Twitter Blue」が実装されたが、日本市場での実装は2023年1月のため、本研究のデータ収集時点では140字制限が適用されていた。

〈発表資料〉

題名	掲載誌・学会名等	発表年月
SNS上のコミュニケーションを探る: Twitter におけるリプライ行動の軽量的分析	第2回計算社会科学学会大会	2023年2月21日