

日本語の大規模 Twitter データからみる

新型コロナウイルスワクチン接種に関する人々の関心の推移

武富有香¹ 須田永遠¹ 中山悠理² 宇野毅明¹ 橋本隆子³

豊田正史⁴ 吉永直樹⁴ 喜連川優^{4,5} Luis E C Rocha^{6,7} 小林亮太^{2,8}

国立情報学研究所 情報学プリンシプル研究系¹ 東京大学大学院 新領域創成科学研究科²

千葉商科大学 商経学部³ 東京大学 生産技術研究所⁴ 国立情報学研究所⁵

ゲント大学 経済学部⁶ ゲント大学 物理・宇宙物理学部⁷ 東京大学 数理・情報教育研究センター⁸

yuka_takedomi@nii.ac.jp sudatowa@nii.ac.jp 3645588575@edu.k.u-tokyo.ac.jp uno@nii.jp

takako@cuc.ac.jp toyoda@tkl.iis.u-tokyo.ac.jp ynaga@iis.u-tokyo.ac.jp

kitsure@tkl.iis.u-tokyo.ac.jp luis.rocha@ugent.be r-koba@edu.k.u-tokyo.ac.jp

概要

日本での新型コロナウイルスワクチン接種は、安全性・有効性に対する不安や接種に関する政策への不満があったにもかかわらず、欧米諸国と比べて迅速に進んだ。短期間で高い接種率に達する過程における人々の考えや関心を知ることは、公衆衛生上の重要な問題である。

本研究では2021年1月から10月につぶやかれた「ワクチン」という語を含む日本語の全ツイートを分析した。トピックモデルを用いてツイートを15のトピックに分類し、意味解釈を行うことにより4つの話題（個人的事柄、ニュース、政治、陰謀論とユーモア）に整理した。そしてテキスト情報を考慮に入れた時系列分析を行った結果、6月の職域接種の開始を境に、政策、有効性、関連ニュースなどのワクチンに関する社会的トピックのツイートの割合が減り、接種の予定や報告、自身の副反応などの個人的事柄に関するツイートの割合が増えたことが明らかになった。

1 はじめに

日本における新型コロナウイルスワクチンの接種開始は、欧米諸国と比べると2ヶ月以上も遅れた。しかしながら、国民の間にワクチンに対する多くの不安や不満があったにもかかわらず、日本におけるワクチン接種は世界の中でも速く進み、医療従事者への接種開始からわずか8ヶ月ほどでワクチン接種率は72%（世界229カ国中14位）に達した。この接種期間中に、人々のワクチンについての興味・関心はどのように変化したのだろうか。

ワクチンについての人々の考え方を調べることは、公衆衛生における重要な課題の1つであり、これまでは主にアンケートによって調査されてきた。しかし、大規模なアンケートを繰り返し実施することには非常にコストがかかる。質問項目の設定によって結果が大きく変わりうることや、被験者が控えめに回答する傾向が多いことなどの問題点も指摘されている。一方、人々の興味や関心を推しはかるために、ソーシャルメディアのデータを分析する研究も行われるようになってきた。ソーシャルメディア上の活動が現実の人々の興味や関心を反映することも先行研究によって示されている[1]。

本研究では、Twitter でつぶやかれた長期間にわたる日本語の大規模データを分析することで、ワクチン接種期間前後にわたる人々の興味や関心の変化を調査した。Twitter や Reddit などのソーシャルメディアを用いて新型コロナウイルスワクチン接種に関する人々の意見を調査する先行研究は存在するが、主に接種期間の初期のみを分析対象としているため、ワクチン接種後の人々の意見や関心の長期的な動向をみることはできていない[2][3]。ひとつの国におけるワクチン接種期間の全ツイートデータを網羅的に分析した研究としては、本研究が最初のものである。

なお、本予稿で発表する内容は論文[4]に基づくものである。解析や結果の詳細に興味を持った読者は当論文を参照されたい。

2 方法

2.1 データについて

本研究のデータセットは NTT データにより提供

された、2021年1月1日から10月31までに投稿された「ワクチン」を含む全ての日本語ツイートを扱った。この期間は日本におけるワクチン先行接種が開始される（2021年2月17日）前の時期から、東京オリンピック・パラリンピックの時期（2021年7月から9月）を経て、日本国民のワクチン接種率が70%に達する（2021年10月25日）までの期間を含んでいる。データについてはツイートのテキスト情報、投稿時間、オリジナルのツイートかリツイートかの情報を併せて取得した。

2.2 データ処理

まず、ツイートからテキスト情報を抽出し、絵文字を除去した。次に、形態素解析エンジン MeCab を使用して品詞ごとに分割し、ストップワード（「これ」「それ」「する」など）を除去した。最後に、各語を原形に変換し（「打た」→「打つ」など）、一人称代名詞（「僕」「おれ」など）などのほぼ同義の語を一つの語（「私」）に統合した。

2.3 トピックモデル

次に、LDA (Latent Dirichlet Allocation) [5]モデルを用いてツイートのトピック（話題）の推定を行った。解析にあたっては、出現頻度が極端に低い語、すなわち、その語の出現するツイート数が1000（全ツイートの0.0004%に相当する）に満たない語、および「ワクチン」と「接種」という出現頻度が極端に高い語は除去した。加えて、bot ツイートを人手によって特定し、除去した。手順としては、LDAによって分類を行い、その結果にbot ツイートのクラスタが出てこなくなるまで、それを繰り返した。トピック数を決めるにあたっては、人間にとっての解釈しやすさを示すとされる Coherence スコア C_v [6] を計算し、最も高いスコアが得られたトピック数15を採用した。

2.4 分割時系列解析

最後に、主要なイベント（東京オリンピックの開幕や職域接種の開始）がツイートの話題に与える影響を評価するために、分割時系列解析 (Interrupted Time Series Regression) [7]を行った。詳細については[4]を参照のこと。

3 結果

2021年1月から10月までの「ワクチン」を含む総ツイート数は114,357,691であった。本研究では引用ツイート、すなわちコメント付きリツイート（5,765,735/114,357,691,全体の5.0%）と言及ツイート（8,416,245/114,357,691,全体の7.4%）は数が少ないため除外し、コメント無しリツイート（75,984,321/114,357,691,全体の66.4%）とオリジナルのツイート（24,191,390/114,357,691,全体の21.2%）のうち後者に着目して分析を行った結果を示す。

3.1 ツイートのクラスタリングと話題

データを収集した10ヶ月におけるツイート群の話題推移を調べるため、トピックモデル (LDA) を用いて、15個のトピックに分類を行なった。各トピックからランダムに抽出したツイート群を実際に読み精査することにより意味解釈を行い、15のトピックに「接種会場からの実況」、「ワクチンの有効性」、「ワクチン政策に関する雑談」などの名前をつけた。これらを4つの主要なテーマ（1. 個人的事柄, 2. ニュース, 3. 政治, 4. 陰謀論とユーモア）に整理した（表1）。

表1 ワクチン関連ツイートの15トピック

テーマ・トピック	ツイート数 (N=989,339) n(%)
1. 個人的な事柄	493,296(49.9)
1.1 ワクチンについての感想	170,095(17.2)
1.2 ワクチン接種の予定	57,763(5.8)
1.3 ワクチン接種会場からの実況	31,952(3.2)
1.4 ワクチン接種の体験記	132,843(13.4)
1.5 ワクチン接種後の痛みや副反応	65,490(6.6)
1.6 ワクチン接種にあたっての準備	35,153(3.6)
2. ニュース	210,550(21.3)
2.1 ワクチンの臨床試験と使用認可	79,247(8.0)
2.2 ワクチンの有効性	74,120(7.5)
2.3 ワクチン接種の予約	57,183(5.8)
3. 政治	169,663(17.1)
3.1 政治に関する意見	95,219(9.6)
3.2 マスメディアの報道に関する意見	41,094(4.2)
3.3 ワクチン政策に関する雑談	33,350(3.4)

4. 陰謀論とユーモア	115,830(11.7)
4.1 人口抑制	41,428(4.2)
4.2 身体への影響	30,221(3.1)
4.3 インターネットミーム	44,181(4.5)

トピック分析を行った結果、最もツイート数が多かったテーマは「個人的な事柄」であり、全体のおよそ50%を占めた。このテーマの中には、接種に対する個人の考え方(1.1)や、いつ接種を受けるかというスケジュールに関する話題(1.2)、「いま会場で並んでいる」「いま打った」など接種体験の実況的な報告(1.3)、接種の体験を日記のように記すもの(1.4)、接種の痛みや発熱などの副反応についての報告(1.5)、ワクチン接種にあたって発熱時に服用する薬や経口補水液などの備えについて記すもの(1.6)が存在した。

2番目に多かったテーマは「ニュース」である。モデルナやファイザー、アストラゼネカなどのワクチン開発について、臨床試験が行われたことや使用認可が下りたことについての日本国内や世界のニュース(2.1)、mRNAワクチンが新型コロナウイルスに対しても有効であることや、アストラゼネカ製ワクチンで血栓などの重大な副作用が起きる可能性や、接種が始まった国で副作用による死者が出たことなど、ワクチンの有効性についてのニュース(2.2)、ワクチン接種の空き状況や予約の仕方などに関する話題(2.3)が存在した。

3番目に多かったテーマは「政治」である。日本の接種開始が他国より遅れをとっていることへの不満や、国民にワクチンが行き渡らない状態で東京オリンピックが開催されることなど、政治に対する意見(3.1)、マスコミが不安を煽っている・マスコミの情報が信じられないといったマスメディアの報道に対する意見(3.3)、ワクチン担当大臣に就任した河野太郎大臣に言及するなど、ワクチン政策の雑談(3.4)がこのテーマに含まれる。

最も少なかったテーマは「陰謀論とユーモア」である。「コロナワクチンの目的は世界の人口を減らすためのものである」といったような人口抑制に関する陰謀論(4.1)、「コロナワクチンは人々を5Gに接続するための策略である」など身体への影響に関する陰謀論(4.2)に加え、いわゆるインターネットミーム(4.3)が存在した。ただし「ワクチンで5Gに接続できるなんてわくわくする!」というような陰謀論を冗談として愉しんでいるツイートが4.2に含

まれるなど、いわゆる陰謀論とみなされるような主張を書いているツイートばかりではない。これらのツイートを含めた「陰謀論」のツイートは7%にすぎなかった。4.3のクラスタには、ワクチンを2回打つことを「ワクワクチンチン」と称したツイートが多数を占め、流行語のような様相をみせた。

3.2 各テーマの時間変化

つぎに、各テーマのツイート割合の時間変化を調べた(図1)。2021年1月当初は、個人的事柄(30%)、ニュース(30%)、政治(25%)、陰謀論とユーモア(15%)というように、4つのテーマにツイートは分散していた。しかし、6月以降には接種後の体調や副反応など個人的事柄に関するツイートが急増し、10月には「ワクチン」を含むツイート全体の70%を占めるに至った。この結果は、6月以降にTwitterユーザーの興味・関心が個人的事柄に関する話題に集中したことを示唆する。

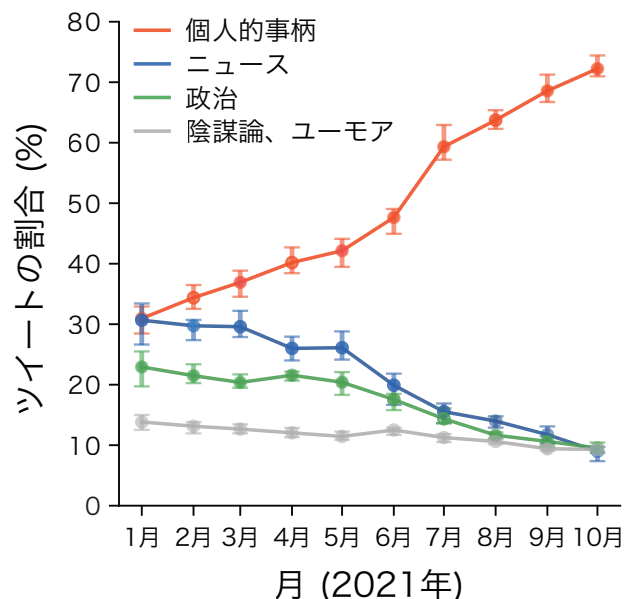


図1. 各テーマのツイート割合の時間変化

3.3 社会的イベントと関心の推移

最後に、4つの社会的イベント(A. 医療従事者の接種開始:2月, B. 高齢者の接種開始:4月, C. 職域接種開始:6月, D. 東京オリンピック:7月)がツイートされる話題に与える影響を中断時系列分析[5]を用いて分析した。この結果、C.職域接種, D.東京オリンピックの2つのイベントがツイートの話題に影響を与えることがわかった。特に、職域接種が

開始された 2021 年 6 月 21 日以降に個人的事柄に関するツイートの割合が急速に増加することが確認された (図 2) . この結果は、多くの Twitter ユーザがワクチン接種を受けることが可能となった職域接種開始以降に、Twitter ユーザの興味・関心が個人的事柄に集中したことを示唆している。

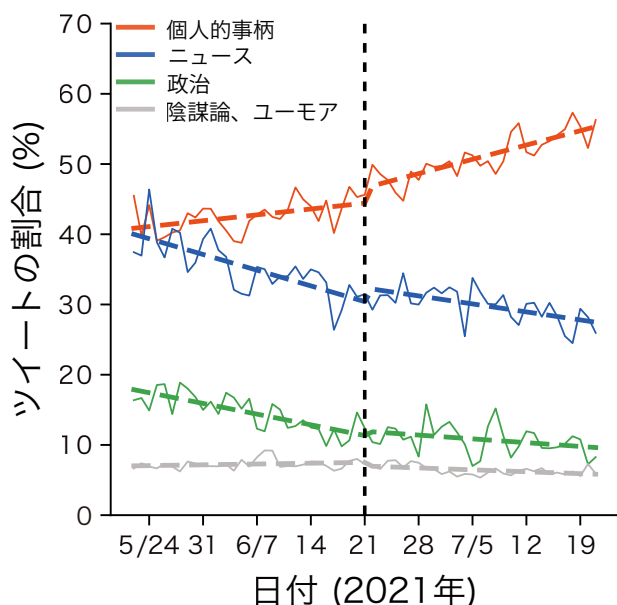


図 2. 職域接種開始前後における各テーマのツイート割合の時間変化

4 結論

本研究では、2021 年 1 月から 10 月までのワクチン接種期間中につぶやかれた「ワクチン」という語を含む日本語の全ツイートを分析した。トピックモデルを用いてツイートを 15 のトピックに分類し、各トピックの意味解釈を行うことにより、4 つの話題 (個人的事柄, ニュース, 政治, 陰謀論とユーモア) に整理した。2021 年 1 月当初、ツイートは 4 つのテーマに分散していたが、6 月以降にはユーザの興味・関心が個人的事柄に関する話題に集中した。さらに中継時系列分析を行った結果、2021 年 6 月の職域接種を境に、ワクチン政策、ワクチン関連ニュースなど社会的な話題に関するツイートの割合が減り、接種の予定や報告、自身の副反応など個人的事柄に関するツイートの割合が増えたことを発見した。

謝辞

本研究は、内閣府 Covid-19 AI・シミュレーションプロジェクトの一環として実施され、科学技術振興機構 (JST) JPMJPR1925, JPMJCR1401, 科学研究費助成事業 基盤研究(A) 19H0113, 21H04571, 基盤研究(B) 21H03559, 22H03695, 基盤研究(C) 18K11560, 22K12285, 二国間交流事業協同研究 JPJSBP1202022 01, 日本医療健康開発機構 JP21wm0525004 による支援を受けて行われました。

参考文献

- [1] Kwak H. et al., (2010). "What is Twitter, a social network or a news media?", WWW '10, pp.591-600.
- [2] Lyu JC, Han EL, Luli GK.,(2011). "COVID-19 vaccine-related discussion on Twitter: Topic modeling and sentiment analysis.", JMed Internet Res 23(6):e24435.
- [3] Wu W, Lyu H, Luo J., (2021). "Characterizing discourse about COVID-19 vaccines: A Reddit version of the pandemic story.", Health Data Science 1-11.
- [4] Kobayashi R, Takedomi Y, Nakayama Y, Suda T et al., (2022). "Evolution of Public Opinion on COVID-19 Vaccination in Japan: Large-Scale Twitter Data Analysis.", J Med Internet Res 24: e41928.
- [5] Blei DM, Ng AY, Jordan ML., (2003). "Latent dirichlet allocation.", J Mach Learn Res 3: pp. 993- 1022.
- [6] Röder M, Both A, Hinneburg A., (2015). "Exploring the Space of Topic Coherence Measures.", In WSDM' 15, pp. 399-408.
- [7] Bernal JL, Cummins S, Gasparrini A., (2017). "Interrupted time series regression for the evaluation of public health interventions: a tutorial.", Int J Epidemiol 46: pp. 348-355.