新型コロナワクチンをめぐるTwitter上の話題変化
テキスト精読と頻出単語分析による仮説構築とその検証

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] 小林亮太[2,6]
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] 情報システム研究機構[5]
東京大学 数理・情報教育研究センター[6]

{yuka_takedomi, sudatowa, uno}@nii.ac.jp yurinakayama620620@gmail.com takako@cuc.ac.jp
{toyoda, kitsure}@tkl.iis.u-tokyo.ac.jp ynaga@iis.u-tokyo.ac.jp r-koba@k.u-tokyo.ac.jp

概要

新型コロナワクチン接種期間中の人々の考えや関心の変化を知るために，「ワクチン」を含む日本語の全ツイート(1.1 億件)を分析した．先行研究では大きな世論のダイナミクスを捉えることに成功したが，本研究ではツイートのより詳細な内容とその時間変化を明らかにすることを目的とした．まず LDA トピックモデルを用いて 15 の主要トピックを特定し，トピックの内容を人手で意味解釈して 4 つの主要なテーマに分類した．次に月毎の頻出単語分析と約 15,000 件のツイート精読を組み合わせ，各トピックの話題変化に関する仮説を構築した．最後に，トピック内での単語の時間変化を調べ，仮説の妥当性を定量的に検証した．その結果，「怖い」という気持ちや強い意見をあらわすツイートは，接種の先行きが最も不明瞭だった時期に多く，その後減少したことや，ワクチン接種を受ける人物が医療従事者，高齢者から自分自身へと変化するのに対応して，ツイートで言及される対象も変化したことが示唆された．

1 はじめに

日本における新型コロナワクチン接種は欧米諸国と比べると 2 ヶ月以上遅れて始まった．日本は先進国の中でもワクチン信頼度が最も低い国として知られており[de Figueiredo 20]，接種期間前にさまざまな不安や不満の空気があったことは記憶に新しい．このような複数の懸念事項がありながら，接種開始から 8 ヶ月後の 2021 年 10 月には，国民のワクチン接種率は 72%（世界 229 カ国中 14 位）に達した．短期間で高い接種率に達した背景にある，人々の心の動きはどのようなものだったのだろうか．特に，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．

多くの先行研究[Lyu 21][Niu 22]は，ワクチン接種開始前後の限られた期間を分析対象としており，接種開始前からワクチンが普及するまでの長期間にわたる人々の意見や関心の変化を観察した研究は少ない．その中で，日本におけるワクチン接種期間のツイートデータを網羅的に分析した小林らの研究[Kobayashi 22]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扱い，トピック分析と意味解釈を行うことによってワクチンに関する世論変化を調べている．その結果，6 月の職域接種の開始を境に，ワクチン政策，有効性，関連ニュースなどのワクチンに関する社会的なトピックに関するツイートの割合が減り，接種の予定や報告，自身の副反応などの個人的な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．この研究は大きな世論のダイナミクスを捉えることに成功しているが，その一方で，社会的なトピックや個人的な事柄に関するツイートで語られていた詳細な内容や，その時間的推移については分析されていない．本研究では，より細かく，人々の考えや関心の時間変化を捉えることを目的とし，先行研究と同じく，2021 年 1 月から 10 月の接種期間前後を含めた期間に投稿された「ワクチン」という語を含む日本語の全ツイート（1.1 億）を分析対象としている．まず，LDA モデルを用いて 15 の主要トピックを特定し，トピックの内容を人手で意味解釈したのちに，15 のトピックを「個人的な事柄」「ニ

ュース」「政治」「陰謀とユーモア」の 4 つのテーマのいずれかに分類した．その上で，頻出単語のリストとツイート本文の人手による精読を通じて，それぞれのトピックで話されていた具体的な内容とその推移に関する仮説を立てた．次に，単語の出現割合の変化を調べることで，得られた仮説の妥当性を調べた．

本予稿の内容は，[武富 24] に基づいており，より詳細な結果は論文に記載されている．本予稿では，個人的な事柄に関する具体的なツイートの内容の時間推移について，仮説と検証の一部を紹介する．本予稿に用いた図表は，上記論文からの引用である．

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 日までに投稿された「ワクチン」を含む全ての日本語ツイート（約 1.1 億件）である．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．なお，オリジナルツイートとはリツイート，メンション (言及ツイート) 以外の通常のツイートを意味する．

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）単語の正規化を行った．

2.3 トピックモデル

本研究では LDA (Latent Dirichlet Allocation) [Blei 03] を用いてツイートのトピック（話題）の推定を行った．LDA を用いた分析を行う前に，1 日に投稿されたツイート数の時間変化を調べた．その結果，7 月から 9 月に多くのツイートが集中していた．本研究ではツイート活動の非定常性を取り除いて各月での変化を調べるために，各月から 10 万ツイートをランダムに抽出した．前処理として，頻度が低い希少語，ここでは登場回数が 1,000 回以下 (出現率 0.0004% 以下) の単語と，最も頻度の高い 2 単語（「ワクチン」，「接種」） を削除した．LDA による分類結果に bot ツイートのクラスタが出てきたため，bot ツイートは人手で特定し除去した．トピック数は，人間にとっての解釈しやすさを示すとされる Coherence スコア CV [Röder 15] 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 トピック内の内容変化の分析

トピックモデルで分類された各トピックのツイートにおける，2021 年 1 月から 10 月の各月における単語の出現頻度を計算し，頻出単語の順位表 (頻出単語リスト) を作成した．つぎに，各月・各トピックについて事後確率の高い順番に抽出した 100 ツイート (合計 15,000 ツイート)を精読することによって意味内容を人手で解釈し，トピック内の話題についての仮説を立てた．そして，対応するトピックの頻出単語リストにおいて，出現頻度の上位 10 位以内に一度以上入った単語の中から，仮説と関連が深そうな単語を選定した．最後に，選定された単語のトピック内における出現割合の時間変化を調べることで，仮説の妥当性を定量的に検証した．

3. 結果

本研究では，該当期間のオリジナルツイート 24,191,390 ツイートを対象とし，各月での変化を調べるため，各月から 10 万ツイートをランダムに抽出したデータセットを分析した．

3.1 ツイートのクラスタリングと話題

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精読して意味解釈を行い，15 のトピックに「接種の体験記」，「ワクチンの有効性」，「ワクチン政策に関する所感」など，トピックを意味づけて名前をつけた．これら 15 のトピックを 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）のいずれかに振り分けた．

表 1. ワクチン関連ツイートから得られた 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 接種に対する個人の考え方	170,095(17.2)
1.2 接種についてのスケジュール	57,763(5.8)
1.3 接種に関する速報	31,952(3.2)
1.4 接種の体験記	132,843(13.4)
1.5 副反応:痛み	65,490(6.6)
1.6 副反応:発熱	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 ワクチン政策についての所感	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)

テーマ 1. 個人的な事柄には，ワクチンに対する意見，接種の予定や接種後の報告などユーザ自身の体験の記述が含まれていた．テーマ 2. ニュースには，ワクチンの使用認可や有効性など国内外のワクチン関連ニュースに関するツイートが含まれていた．テーマ 3. 政治には，政策やマスメディアに対する意見が含まれていた．テーマ 4. 陰謀論とユーモアには，陰謀論について言及しているものや，冗談やユーモアが含まれていた．

これら 4 つのテーマの割合の変化を分析した先行研究 [Kobayashi 22] では，2021 年 1 月に同程度であったテーマ 1, 2, 3 の割合が 6 月以降に急激に変化し，テーマ 1 の接種の予定や報告，副反応など個人的な事柄に関するツイートの割合が急増したことが明らかになった．この結果は，職域接種の開始を境にして Twitter ユーザの興味，関心が個人的な事柄に集中したことを示唆している．

3.2 トピックの詳細な内容と話題の推移

15 のトピック内における内容の時間変化を調べるため，頻出単語リストとツイート本文を手がかりに仮説を立て，仮説の妥当性を検証した．本予稿では全ての結果を詳述できないため，特筆すべきトピックについて構築した仮説とその検証の結果を記す．

トピック分析の結果，最もツイート数が多かったテーマである「個人的な事柄」（全体の約 50%） には，接種期間中の人々の考えや行動，関心の記述が含まれており，注目すべき点が多い．

ワクチンについての個人的な考え（トピック 1.1）では，ワクチン接種をめぐる個人の意見や副反応のリスクについての感情が表明されていた．代表的なツイートを精読したところ，一人称単数を用いて意見や感情の表明を行うツイートが多く見られた（ツイートの例: 「ワクチンかぁ... 打つ？打たない？私は打つつもり」，「私はコロナよりコロナのワクチンを打つことが怖いけど，上司は早く打ちたいんだって」）．日本語では通常，一人称の主語は省略されることが多い．しかし，このように”（周りの人はどう考えているかはわからないが）私は（このように）思う”というように，特に強調する場合には一人称が用いられるとされる [成山 09]. 以上から，

仮説 1 「私」という一人称単数を用いて自分自身の意見，意思や感情を表明したツイートはワクチン接種をめぐる意見が多様な時期に多くあらわれ，事態が落ち着くにつれて減っていく．

を立てた．本トピックで頻出していた単語の出現割合を調べたところ，「私」の出現割合は 1 月が最も高く，その後減少したため（**図 1A**），仮説 1 は支持されたといえる．さらに，

仮説 2 「思う」を用いて自分の意見を表明したツイートは減っていく．

を立てた．しかし，「思う」の出現割合は大きく変動しなかったため（**図 1A**），仮説 2 は支持されなかった．この結果から，個人の意見の表明は継続して行われていたと推察される．

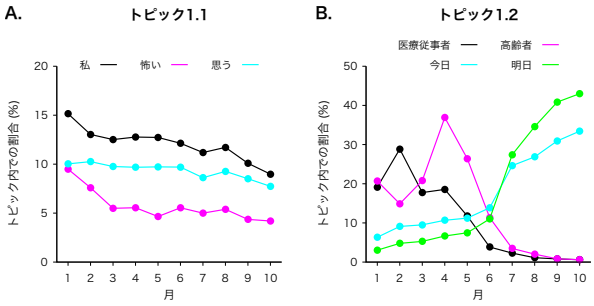

図 1　個人的な事柄(テーマ 1) の 2 つのトピック（A. 接種に対する個人の考え方(トピック 1.1)，B. 接種についてのスケジュール(トピック 1.2)）における頻出単語の出現割合の推移.
2021 年 1 月から 10 月までの各月における，代表的な頻出単語の各トピック内での出現割合(%)が示されている.

また，「怖い」は 1 月，2 月に頻出していたことから

仮説 3　「怖い」という感情表現は徐々に減っていく.

を立てた．「怖い」の出現割合は 1 月が最も高く，その後減少していたため，結果は仮説 3 を支持していた．多くの人はまだ接種する段階に至っていないものの，医療従事者の先行接種により徐々にワクチン接種についての知見が増えていったことや，深刻な副反応による大きな混乱がなかったことも「怖い」という気持ちの表明の減少に影響したのではないかと思われる.

ワクチン接種についての個人の予定（トピック 1.2）では，1 月から 5 月には優先接種のスケジュールについてのツイートが多くみられ（例:「今日から大阪でも医療従事者のワクチン先行接種が始まった」），7 月以降にはいつ・何回目のワクチンを受けるかという自分自身の予定に関するツイートが多くみられた（例:「明日ワクチン 2 回目！」）．このことから，

仮説 4 ツイートされた時にワクチンを受けていた人物が言及される傾向にある.

を立てた．頻出単語の出現割合の変化を調べると，「医療従事者」は 2 月，「高齢者」は 4 月に最大になっており，6 月以降には「明日」，「今日」が増加していた（**図 1B**）．このことは，2 月に医療従事者対象，4 月に高齢者対象のワクチン接種が開始され，6 月に職域接種が開始されたことと対応していると思われる．また「今日」，「明日」を含む代表的ツイートを確認したところ，「明日ワクチン 2 回目！」「今日ワクチン打つ」など，ユーザ自身の接種予定の報告が，一人称を省略して書かれていた．以上から，ワクチン接種を受ける人物が医療従事者，高齢者から自分自身へと変化するとともに，ツイートで言及された対象も変化したことが推察され，仮説 4 は支持される.

4 おわりに

本研究では，新型コロナワクチン接種期間中の人々の考えや関心の変化を知るために，「ワクチン」を含む日本語の全ツイートを分析対象とし，投稿内容の詳細とその時間変化を分析した．まず LDA トピックモデルを用いて 15 の主要トピックを特定し，トピックの内容を人手で意味解釈し，さらに 4 つの主要なテーマに分類した．次に月毎の頻出単語分析と約 15,000 件のツイート精読を組み合わせ，各トピックの話題変化に関する仮説を構築した．最後に，各トピック内での単語の時間変化を調べ，仮説の妥当性を定量的に検証した.

本予稿では，個人的な事柄についてのツイートに関する分析を紹介した．ワクチンについての個人的な考え（トピック 1.1）については，「私」という一人称単数を使用して意見表明をするツイートはワクチン接種の先行きが最も不明瞭だった時期に多く，事態が落ち着いていくにつれて減少していったこと，意見や気持ちの表明自体は継続して行われていたこと，「怖い」という感情表現は 1 月に最も割合が高く，医療従事者の接種開始 (2 月) によってワクチン接種を国内で実際に受けた人があらわれるにつれて減少していたことがわかった．また，ワクチン接種についての個人の予定（トピック 1.2）については，ツイートされた時期にワクチンを受けていた人物がツイート内で言及されており，ワクチン接種を受ける人物が医療従事者，高齢者から自分自身へと変化するのに対応して，ツイートで言及された対象も変化していることがわかった.

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）JPMJCR1401，科学研究費助成事業 基盤研究(A) 19H01133，21H04571，基盤研究(B) 21H03559,22H03695,基盤研究(C)18K11560,22K12285,24K15204，日本医療健康開発機構（AMED）JP21wm0525004，JP223fa627001 による支援を受けて行われた.

参考文献

[de Figueiredo 20] de Figueiredo, A, Simas, C, Karafillakis, E, Paterson, P, Larson, HJ: Mapping global trends in vaccine confidence and investigating barriers to vaccine uptake: a large-scale retrospective temporal modelling study, Lancet 2020, 396, pp.898-908 (2020)

[Lyu 21] Lyu JC, Han EL, Luli GK: COVID-19 vaccine-related discussion on Twitter: Topic modeling and sentiment analysis, J Med Internet Res 23: e24435 (2021)

[Niu 22] Niu Q, Liu J, Kato M, Shinohara Y, Matsumura N, Aoyama T, Nagai-Tanima M: Public opinion and sentiment before and at the beginning of COVID-19 vaccinations in Japan: Twitter Analysis, JMIR Infodemiology 2: e32335 (2022)

[Kobayashi 22] Kobayashi R, Takedomi Y, Nakayama Y, Suda T, Uno T, Hashimoto T, Toyoda M, Yoshinaga N, Kitsuregawa M, Rocha LEC: Evolution of public opinion on COVID-19 vaccination in Japan: Large-scale Twitter data Analysis, J Med Internet Res 24: e41928 (2022)

[Blei 03] Blei DM, Ng AY, Jordan MI: Latent Dirichlet allocation, J Mach Learn Res 3: pp.993-1022 (2003)

[Röder 15] Röder M, Both A, Hinneburg A: Exploring the space of topic coherence measures, Proceedings of the 8th ACM International Conference on Web Search and Data Mining: pp.399-408 (2015)

[武富 24] 武富有香，中山悠理，須田永遠，宇野毅明，橋本隆子，豊田正史，吉永直樹，喜連川優，小林亮太. Twitter における新型コロナワクチンに関する話題の変化 ツイート本文の読解を通じた仮説構築による分析. 人工知能学会論文誌，2024，39 巻，5 号，p.C-N93_1-10.

[成山 09] 成山重子，日本語の省略がわかる本誰が？誰に？何を？，明治書院（2009）