

Twitter データを用いた新型コロナワクチン 賛成派と反対派の傾向分析

田中 創太†

細部 博史‡

法政大学情報科学部†

法政大学情報科学部‡

1. はじめに

日本国内で初めて新型コロナウイルスへの感染が確認されてから現在まで約 3 年間、感染者数などのデータ、政府の対応、ワクチンの安全性など様々な話題で議論が繰り返されてきた。この議論の多くは、新型コロナへの感染を避けワクチンを自ら進んで接種する賛成派と、新型コロナ自体の存在を否定しワクチンは有害だと主張する反対派の間で行われてきた。反対派は少数ではあるが SNS 等で進んで情報発信する人が多く、反対派の存在が国民の不安を煽り集団免疫の形成を遅らせている要因であると考えられる。

本研究は、コロナに関するツイートの分析をした鳥海ら [1]の研究と Baj-Rogowska [2]の研究を参考に、ワクチンの賛成派と反対派の概念を加えることで、さらに実用的なデータを得て反対派の増加を抑えるための戦略を考える助けとすることを目的とする。具体的には、(1)コロナかワクチンに関するツイートをしたユーザー全体のネットワーク内に存在する賛成派と反対派について、人数分布の差と時系列による変化を見るユーザーの偏り分析、(2)賛成派と反対派のツイート全文から感情を抽出し、その種類と数の差とその時系列による変化を見る感情分析、(3)賛成派と反対派それぞれが注目している話題の差とその時系列による変化を見るトピック分析の三つを行う。

2. 提案手法

2.1. 傾向分析

新型コロナウイルスワクチンの賛成派と反対派の傾向分析を行うための方法を提案する。その流れは、(1)ツイートデータの収集、(2)リツイートデータの収集、(3)賛成派と反対派の区別、(4)ユーザーの偏り分析、(5)感情分析、(6)トピック分析である。

データの収集と賛成派と反対派の区別を行った後、三つの分析を行う。まずユーザーの偏り分析でそれぞれの人数やコミュニティの形成状況を大まかに見る。その後、感情分析とトピック分析で細かい傾向を見る。これらの結果を併せて見ることでより詳細で正確な分析を行う。

2.2. 賛成派と反対派の区別

賛成派と反対派はツイート本文に含まれる単語によって区別する。例を挙げると、ツイートの中に「ワクチン」と「有害」を同時に含んだ場合はこのユーザーを反対派

であると区別する。本研究では国民の不安を煽るような人を反対派とするため、ワクチンを接種するメリットを理解しつつも接種しない選択をした人は反対派とせず賛成派とする。賛成派の範囲はかなり広く、単語で絞り込むことが難しいため、指定した単語を含まないツイートはすべて賛成派であるとする。

本研究で指定した単語は、ワクチン信者、ワク信、コロナ脳、兵器、茶番、有害、ワクチン臭、人殺し、殺人、毒（消毒、毒性、弱毒、強毒、気の毒を除く）、マスクはずそう、マスクを外そう、コロナ劇場、奴隷、ワクチン中止、薬害、普通の風邪、ただの風邪、梓珍の 19 個である。

2.3. ユーザーの偏り分析

ユーザーの偏り分析のうち、ネットワークの構築、コミュニティの抽出までは鳥海ら [1]の研究を参考に行う。結果の示し方について、鳥海らは KL 情報量を用いていたが、本研究では偏りだけでなく賛成派と反対派の人数の差や偏り方も結果として示すため、Fruchterman-Reingold 法を用いたグラフ描画で結果を視覚的に示す。これによって人数の差や偏り方を直観的に把握できるようにする。

3. 実験

3.1. データの概要

本研究の実験に用いたツイートデータ、リツイートデータの概要を以下に示す。

3.1.1. ツイートデータ

収集したデータの条件は「コロナ」「ワクチン」のいずれかを含むツイートであること（ただし、Twitter に承認されたユーザーのツイートは除く）で、時系列による変化を見るため投稿日時を 2021 年 8 月、12 月、2022 年 4 月、8 月の四つの月に絞った。また、平日と祝日で国民の生活が変化し、Twitter 上での振る舞いも変化する可能性を考慮し、収集したツイートの投稿日時は上記の各月の第 1 日曜日と第 1 水曜日の二つの日で合計八つの日を指定している。データ数は各日約 24,000 件である。収集するデータの内容は、ツイート ID、ツイートをしたユーザーの ID、ツイート本文の三つである。

3.1.2. リツイートデータ

収集したツイートに対してリツイートしたユーザーの ID を収集し、ツイートをしたユーザーの ID とともに保存する。リツイートデータの件数は、最も少ない 2022 年 8 月 7 日で 12,307 件で、最も多い 2021 年 12 月 1 日で 18,620 件である。

3.2. ユーザーの偏り分析

収集したデータからネットワークの作成、コミュニティの抽出をした後、グラフの描画を行った。2021 年 8 月

Tendency Analysis of Proponents and Opponents of COVID-19 Vaccines Using Twitter Data

†Sota Tanaka, Faculty of Computer and Information Sciences, Hosei University

‡Hiroshi Hosobe, Faculty of Computer and Information Sciences, Hosei University

の結果を図1に示す。時系列順に見ると、賛成派のグラフは大きな変化がないが、反対派のグラフからはその勢力は時が経つにつれて拡大し、密なコミュニティを形成し、賛成派のグラフの形に近づいていく傾向にあることが見て取れる。

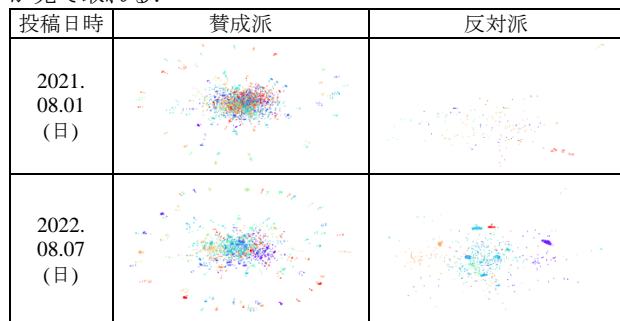


図1 ユーザーの偏り分析

3.3. 感情分析

収集したツイートの文章に対して ML-Ask で感情分析を行い、10種類の感情がどのくらい割り当てられたのかを賛成派、反対派に分けて時系列順に求めた結果を図2に示す。賛成派、反対派ともに嫌の感情が最も多くの割合を占めていることが分かる。賛成派と反対派を比較すると、反対派は安と喜の感情が少なく、怖と嫌の感情が多くなっていることが分かる。

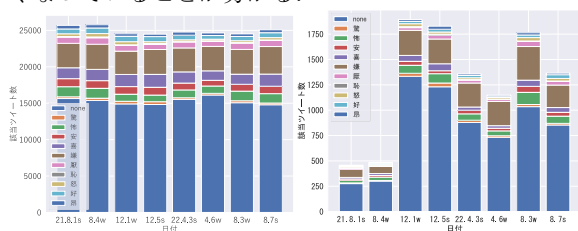


図2 感情分析結果 (左: 賛成派, 右: 反対派)

3.4. トピック分析

ツイートデータを用いて LDA で学習を行い、賛成派と反対派それぞれのトピックを抽出した。賛成派ではすべての時系列において自分や周りの状況、ニュースに対する感想が大半を占め、コロナ禍やワクチンに対する個人的な意見はほとんどないことが分かった。反対派では当初からワクチンが危険だというトピックが多いのではなく、時間が経つにつれてワクチンは危険だと発信するようになっていくことが分かった。また、反対派は賛成派と比べて細かいデータまで目を通しその真偽を疑っている様子が多く見られた。

3.5. 評価

賛成派と反対派を区別する手法の評価をするためにツイート本文から協力者3名にそれぞれ賛成派と反対派を分けてもらい、それを正解データとして提案手法の精度を測る実験を行った。実験に用いたツイートデータの数は240件である。

結果を表1に示す。賛成派に関してはかなり高い正解率だが、反対派に関してはあまり高くない正解率だった。また、リツイート(RT)関係を除く反対派、つまり単語による区別のみ行った場合の反対派に関する正解率の方が反対派より少し高かった。

表1 提案手法の評価

	賛成派	反対派	全体	反対派(RT関係除く)
正解率	96.6%	61.3%	79.1%	64.1%

4. 議論

提案手法の精度は賛成派に対しては良かったが、反対派に対しては高くならなかった。最大の原因として、反対派に対して否定的な意見を述べているツイートも指定した単語を含んでいることが多いことが考えられる。リツイート関係を含まない方が高い正解率だった原因も、単語による分類の正解率が低いためだと考えられる。この問題の解決策として、賛成派と反対派による議論がすべて反対派に分類されることを防ぐためにリプライの除外や、指定した単語に対して「なんか」「なんて」「じゃない」といった否定的な単語を並べているツイートの除外が挙げられる。

偏り分析から、反対派は時間が経つにつれて勢力を拡大させ、密なコミュニティを作ることが分かった。トピック分析の2021年8月の部分を見ると、当初から反対派である人は政治への関心が高く、政府への不満も高いことが分かる。つまり、政府への不信感により、政府が出した情報は信頼できないという感情からコロナは存在しない、ワクチンは危険だという発想が生まれ、それが改善されないコロナ禍や、効果が不確実なワクチンに対して不満を持つ人に広がって、反対派の勢力を拡大することにつながっていると考えられる。

感情分析から反対派は否定的な感情が占める割合が多いことが読み取れるが、それ以外で賛成派と反対派の差や時系列による変化はあまり読み取れず、感情を割り当てられなかったツイートが大半を占めた。結果として、本研究の手法で感情分析から望ましい結果を得ることはできなかった。

5. おわりに

ツイートデータから新型コロナウイルスワクチンの賛成派と反対派を分類し、それぞれの傾向について三つの分析を行った。偏り分析とトピック分析では賛成派と反対派の間に明確な差が見受けられたが、感情分析においては大きな差が見られなかった。

反対派に対する分類の精度の低さが大きな問題である。指定する単語の見直しやリプライを除外するなど、指定した単語を使用した賛成派を反対派に分類しなくするなどの改善が必要である。

文献

- [1] 鳥海不二夫, 榎剛史, 吉田光男, "ソーシャルメディアを用いた新型コロナ禍における感情変化の分析," 人工知能学会論文誌, vol. 35, no. 4, pp. F-K45_1-7, 2020.
- [2] A. Baj-Rogowska, "Mapping of the Covid-19 Vaccine Uptake Determinants From Mining Twitter Data," *IEEE Access*, vol. 9, pp. 134929-134944, 2021.