

令和2年度繰越 厚生労働行政推進調査事業補助金(新興・再興感染症及び予防接種政策推進研究事業)
「マスキング時や新興・再興感染症の発生に備えた感染症サーベイランスの強化と
リスクアセスメントに関する研究」

データ分析
(症状ツイート可視化システムの構築と新型コロナウイルス感染症に対する市の反応調査)

研究分担者

荒牧英治 奈良先端科学技術大学院大学・先端科学技術研究科・教授

研究協力者

若宮 翔子 奈良先端科学技術大学院大学・先端科学技術研究科・准教授

勘場 大 奈良先端科学技術大学院大学・先端科学技術研究科・大学院生

上原 誠 奈良先端科学技術大学院大学・先端科学技術研究科・大学院生

研究要旨

感染症発生に関する一般市民の反応を大規模にモニタリングするために、感染症に関わる主な 8 つの症状(熱、頭痛、風邪、悪寒、悪心、下痢、発疹、痙攣)のソーシャルメディアデータ(ツイート)を収集して、ツイート時系列とツイート中の話題語を可視化するウェブベースシステムを構築した。提案システムにより、ツイート時系列のトレンドの短・中・長期的なモニタリングを継続している。また、症状だけでなく、フェイクニュースやデマに関するツイートを対象としたウェブベースシステムも構築した。さらに、市民の新型コロナへの関心、特に不満、を可視化する指標として、「コロナのせい」という句の頻度に着目した市民の反応を分析した。

A. 研究目的

マスキング時や新興・再興感染症の発生に備え、インターネット上で一般市民の反応を大規模にモニタリングする。さまざまな感染症の発生に備えるため、感染症そのものではなく、感染症に関わる主な症状をベースに、ソーシャルメディアデータを収集する。

そして収集したソーシャルメディアデータを分類した結果を定量的・定性的に可視化するウェブベースシステムを構築する。

さらにソーシャルメディアデータを通して、新型コロナウイルスに対する市民の反応をモニタリングする。

B. 研究方法

ソーシャルメディアとして日本でも利用者が多い Twitter を対象に、ツイートデータを収集する。感染症に関わる主な症状として、次の8つの症状を用いる: 熱、頭痛、風邪、悪寒、悪心、下痢、発疹、痙攣。

これらのデータの収集は、奈良先端科学技術大学院大学ソーシャル・コンピューティング研究室において2016年5月から行っており、現在も稼働中である。

また、ソーシャルメディアにおけるフェイクニュースの話題の変動をモニタリングするための可視化システムを構築した。具体的には、Twitter 上でデマを指摘し

ているツイートを収集し、1週間分のツイートテキストをクラスタリングし、分類されたグループ(クラスター)ごとのツイート数の推移や、各クラスターにおけるツイートテキストやユーザプロフィールテキストの特徴語を可視化した。

また、市民の新型コロナへの関心、特に不満、を可視化する指標として、「コロナのせい」という単語の出現頻度を分析している。具体的には、2020年1月から2021年11月30日における「コロナのせい」を含む日本語ツイートを対象とした。「コロナのせい」に後続する文脈には、「(動詞)した」「(動詞)できなかった」が頻回に現れる。そのため、この動詞部分を集計することで、新型コロナの流行に伴い、本来するはずではなかったが、新たに必要に迫られたこと、また、本来したかったが、できなかったことの収集を行なっている。後者については、行われなかったことであるため、他の行動モニタリング指標での検出が困難であるが、本指標により、このような人々の行わなかった行動を計測が可能となる。

(倫理面への配慮) 本研究で収集し解析するソーシャルメディアデータは、原則としてインターネット上に公開されているデータのみを対象としている。

C. 研究結果

図1にデマ言及ツイート可視化システムのイメージを示す。この可視化システムを通して、一部のユーザの繰り返しの投稿によって引き起こされた短命な話題や、世界的なニュースに伴って推移する大きな話題など、デバンキングされる話題の多様性を観察することができた。

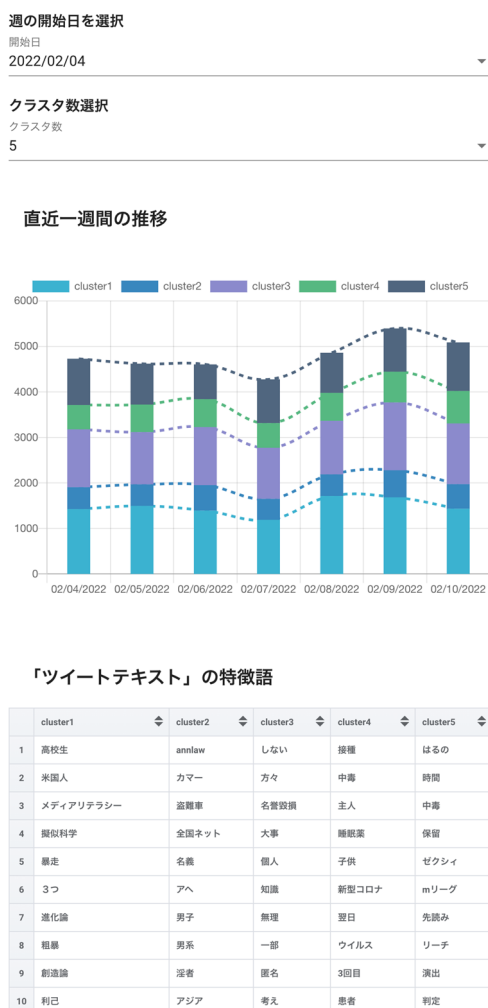


図1. デマ言及ツイート可視化システムのイメージ。分析する週を開始日を選択すると、直近一週間におけるツイートテキストをクラスタリングした結果、クラスターごとのツイート数の推移や特徴語などが提示される。

図2に「コロナのせい」ツイート割合とコロナの陽性患者数の推移を示す。「コロナのせい」ツイートにおいて高頻度で現れた名詞と動詞(いずれも原形に変換して集計)はそれぞれ下記の通りであった。

名詞:「卒業式」「中止」「学校」「仕事」「イベント」「ライブ」「トイレトペーパー」「バイト」「休校」「友達」

動詞:「行く」「なくなる」「無くなる」「休む」「行ける<否定>」「終わる」「出る」「見る」「頑張る」「買う」

ただし、「コロナ」はクエリに含まれている単語であるため上記からは除外した。名詞では時制を表すワードを除外すると、「仕事」「中止」「家」「ライブ」の頻度が高かった。動詞では「行く」が最も頻度が高かったが、実際のツイートでは「行けなかった」や「行かなかった」のように否定系で使われることや、「行きたかった」のように本来「行く」はずであったが実現しなかった場合も多く含まれていた。

D. 考察

図2に示したツイート率と陽性患者数の推移より、コロナウイルスの初期の観測時点や初回の緊急事態宣言前でツイートの割合が多くなり、国民の関心の高さが伺える。ツイート率が最も高くなった2020年2月28日に政府は全国の小学校、中学校、高等学校の一斉休校を発表した。この日の最頻名詞と動詞には、それに関連する「卒業式」「中止」「無くなる」「休む」「行ける<否定>」など一斉休校に関連する単語が見られた。また、トイレトペーパーが無くなるというデマの拡散に関連する単語も確認された。第1回緊急事態宣言中において、感染者の減少に伴い、ツイート割合も減少傾向を示した。2020年6月～9月ごろの第2波においては、感染者の増減に伴い、ツイートも同様の傾向を示した。第3波以降は、第1波や第2波に比べて感染者数は増加しているものの、ツイート割合には大きな変動はなかった。全体的な傾向として、陽性者数に関わらず「コロナのせい」のツイート率は減少傾向にあり、コロナにより影響を受ける生活が人々にとって当たり前になっていることが示唆される。「コロナのせい」と高頻度で共起した動詞の多くは「行く」「見る」「会う」など人と接するものであり、肯定表現であっても否定的な意味で用いられる場合が多かったため、自粛と整合していることがわかった。

E. 結論

従来の症状ベースのモニタリングに加え、フェイクニュースやデマに関するツイートの話題の推移の可視化や、「コロナのせい」に関する解析を加えることで、国民が行えなかった行動の調査を行った。これらにより、話題の継続性や深刻度を考慮したフェイクニュースへの対応の検討や、従来では測定が困難な心的状態

のモニタリングが可能となる可能性があり、リスクコミュニケーションの一助になると期待される。

F. 研究発表

1. 論文発表

1. Taichi Murayama, Nobuyuki Shimizu, Sumiko Fujita, Shoko Wakamiya, Eiji Aramaki: Predicting regional influenza epidemics with uncertainty estimation using commuting data in Japan, PLOS ONE 16(4):e0250417, 2021

2. Taichi Murayama, Shoko Wakamiya, Eiji Aramaki, Ryota Kobayashi: Modeling the spread of fake news on Twitter. PLOS ONE 16(4):e0250419, 2021

2. 学会発表

1. 上原誠、久田祥平、若宮翔子、荒牧英治：Debunking Visualizer: ユーザ情報を考慮したデマ言及ツイート可視化システム、2022年度 人工知能学会全国大会(第36回)

2. Taichi Murayama, Shoko Wakamiya, Eiji Aramaki: Mitigation of Diachronic Bias in Fake News Detection Dataset, In Proceedings of the 7th Workshop on Noisy User generated Text (W-NUT)

G. 知的財産権の出願・登録状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

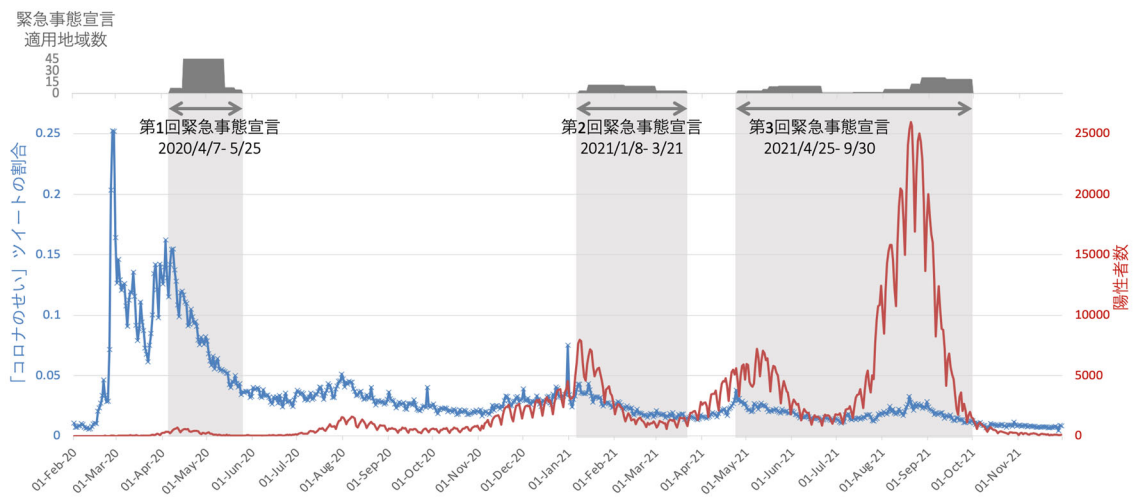


図 2. 「コロナのせい」ツイート頻度(青線)と陽性者数(赤線)の推移。分析期間中に適用された3回の緊急事態宣言の期間を灰色塗り潰して示す。また、各緊急事態宣言中、適用地域の変更があったため、緊急事態宣言が発令された地域数をグラフ上部の棒グラフにより示す。