

分担研究報告書

症状ツイート可視化システムの構築と新型コロナに対する市民の反応調査

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）

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、言語処理により症状と関係するもの（陽性ツイート）と症状とは関係しないもの（陰性ツイート）を分類し、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が可能となっている。今年度は、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、「コロナのせいで」という句の頻度に着目し、市民の反応のモニタリングを試みた。

A. 研究目的

マスクギャザリング時や新興・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、新型コロナウイルス感染症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。表1にこれら8つの症状に関するツイートを収集するために用いたキーワードを示す。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。

本年度はこれらに加えて、新型コロナに関するネガティブな関心の高さを示す「コロナのせい」という単語の出現頻度に注目する。「コロナのせい」に後続する文脈は、「(動詞)した」「(動詞)できなかった」が頻回に現れ、この動詞部を集計することで、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が期待される。特に、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。例えば、新型コロナによるテレワークなどの働き方の変化は、新たに購入したオンライン会議用のカメラの販売実績をみれば分かるが、新型コロナのために、行けなかった海外旅行に

ついでの実測は困難である。本指標は、このような人々の行わなかった行動を計測できる点が新しい。

(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Twitterが公開している公式APIを用いて取得可能な、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。また、今回構築しているシステム上では、ツイート数の時系列とツイートを分析して抽出された特徴語のみを提示している。

C. 研究結果

症状ツイート可視化システムのユーザインターフェースを図1に示す。症状ごとに、ツイートの時系列を可視化したグラフと、ツイートから抽出された話題語が提示される。時系列は陽性ツイートと陰性ツイートに分類した結果を積み重ねて表示される。可視化する時系列の期間は切り替え可能となっており、現在は、1週間、1ヵ月、6ヶ月、1年間、2年間、そして3年間まで、短・中・長期的な時系列のトレンドをそれぞれ確認することができる。図2に「コロナのせい」ツイート頻度の時間変化を示す。12月末に「コロナのせい」という単語自体の出現頻度が高まっており、「クリスマス」、「年始」、「ライブ」などの単語と共起している。

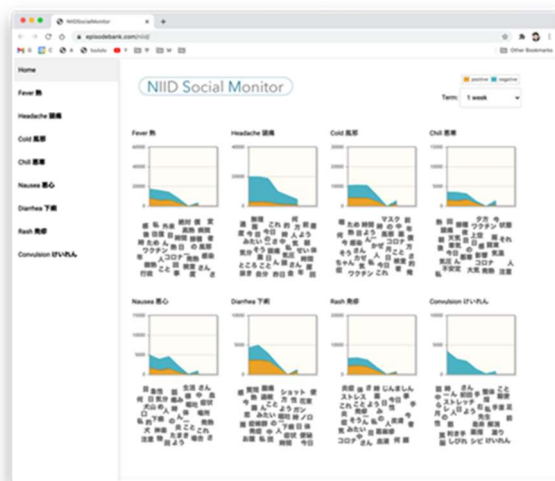


図1. 症状ツイート可視化システムのユーザインターフェース。症状ごとに、ツイートの時系列を可視化したグラフ（横軸が時間、縦軸がツイート数、黄色の塗りつぶしが陽性ツイート数、青色の塗りつぶしが陰性ツイート数）とグラフ下部がツイートから抽出された話題語群。

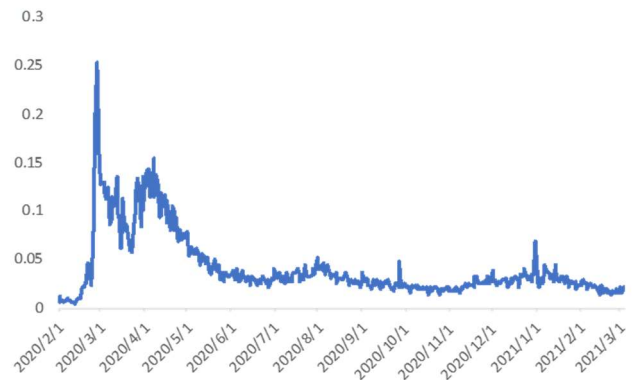


図2: 「コロナのせい」ツイート頻度の時間変化。12月末に「コロナのせい」という単語自体の出現変化が高くなっており、「クリスマス」、「年始」、「ライブ」などの単語と共起している。

D. 考察

従来の症状ベースのモニタリングに加え、「コロナのせい」に関する解析を加えることで、市民が行えなかった行動が把握できる。例えば、実際の患者数はそれほど多くなかった年末における市民の予定変更は大きく、大きなストレスかかっていたものと想像される。特に頻度が上位であった「クリスマス」や「ライブ」など年末特有の行事への期待が高かったことが伺いられる。一方、「帰省」「正月」といった単語は上位には存在しない。

このように、工夫したキーワード「コロナのせい」などを用いることで、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。

E. 結論

感染症に関わる主な症状や新型コロナのせいで抑制された行動を分類して可視化するウェブベースシステムの構築を行い、一般市民の話題を把握可能なツールとした。来年度も継続して運用し、一般市民の反応をモニタリングするとともに、システムの評価を実施する予定である。症状ツイート可視化システム

: <https://www.episodebank.com/nid/>

F. 研究発表 (発表雑誌名巻号・頁・発行年等も記入)

1. 論文発表

1) Murayama T, Shimizu N, Fujita S, Wakamiya S, Aramaki E, Kobayashi R: Modling the spread of fae news on Twitter, PLOS ONE, 2021 (採録決定)

2) Murayama T, Shimizu N, Fujita S, Wakamiya S, Aramaki E: Predicting Regional Influenza Epidemics with Uncertainty Estimation using Commuting Data in Japan, PLOS ONE 2021 (採録決定)

2. 学会発表

(査読あり)

Taichi Murayama, Shoko Wakamiya, and Eiji Aramaki: Fake News Detection using Temporal Feacures Extracted via Point

Process, In Proceedings of the International Workshop on Cyber Social Threats (CySoc2020), 2020

(査読なし)

- 1) 村山太一、若宮翔子、荒牧英治: フェイクニュース検出データセットにおける通時的バイアス、言語処理学会第27回年次大会 (NLP2021)、2021 (2921/3/7、オンライン)
- 2) 村山太一、若宮翔子、荒牧英治、小林亮太: Twitterにおけるフェイクニュース拡散モデルの提案、第13回データ工学と情報マネジメントに関するフォーラム (DEIM2021)、2021 (2021/3/3、オンライン)

G. 知的財産権の出願・登録状況(予定を含む)

1. 特許取得
なし
2. 実用新案登録
なし
3. その他
なし

H. 参考文献

なし

表1: ツイート収集に用いたキーワード

症状	キーワード
熱	高熱、微熱、発熱、熱が高い、熱があった、熱がある、熱が出た、熱風邪
頭痛	頭痛、がんがん、ガンガン
風邪	風邪
悪寒	寒気、悪寒、さむけ
悪心	吐、嘔吐、おう吐、気分悪
下痢	痢
発疹	痒、ブツブツ、蕁麻疹、発疹、ジンマシン、じんましん
痙攣	シビレ、痙、攣