

症状ツイート可視化システムの構築

研究分担者 荒牧 英治 奈良先端科学技術大学院大学 研究推進機構 特任准教授

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を可能にすることで、一般市民の反応の通常状態を知るとともに、異常状態を迅速に検知し、話題の変化を捉えるための一助になると期待される。

A. 研究目的

マス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の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。

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。

キーワードで収集したツイートの中には、症状とは関係のないツイートも含まれている。そのため、発信者が自分自身やごく身近な人の症状について述べているツイート（陽性ツイート；positive）

と、そうでないツイート（陰性ツイート；negative）を自然言語処理により自動的に分類する。分類したツイートを症状ごとに日単位で集計し、時系列を可視化する。また、症状ごとにどのような話題が盛り上がっているのかを直感的に把握するために、各症状のツイートにおける話題語（特定の症状のツイートでは頻出するが、他の症状のツイートでは頻出しない）を抽出し、タグクラウドを作成する。

（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ヶ月、

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。

D. 考察

各症状の3年間のツイートの時系列のトレンドについて考察する。「熱」「悪寒」のツイートの時系列には季節性が顕著に現れており、毎年冬に盛り上がりを示すことが確認された。「熱」「悪寒」ほど顕著ではないものの、「下痢」は冬と夏に、「発疹」は夏に、「悪心」は冬に、周期的な盛り上がりを示すことが確認された。「風邪」「頭痛」「痙攣」については、キーワードを含むツイートはあるものの、そのほとんどが陰性ツイートであることが確認された。このように、時系列のトレンドの短・中・長期的なモニタリングを可能にすることで、一般市民の反応の通常状態を知るとともに、異常状態を迅速に検知し、話題の変化を捉えるため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のソーシャルメディア

データ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て公開した。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/niid/>

F. 研究発表

1. 論文発表
なし
2. 学会発表
なし

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

表 1. 8つの症状に関する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

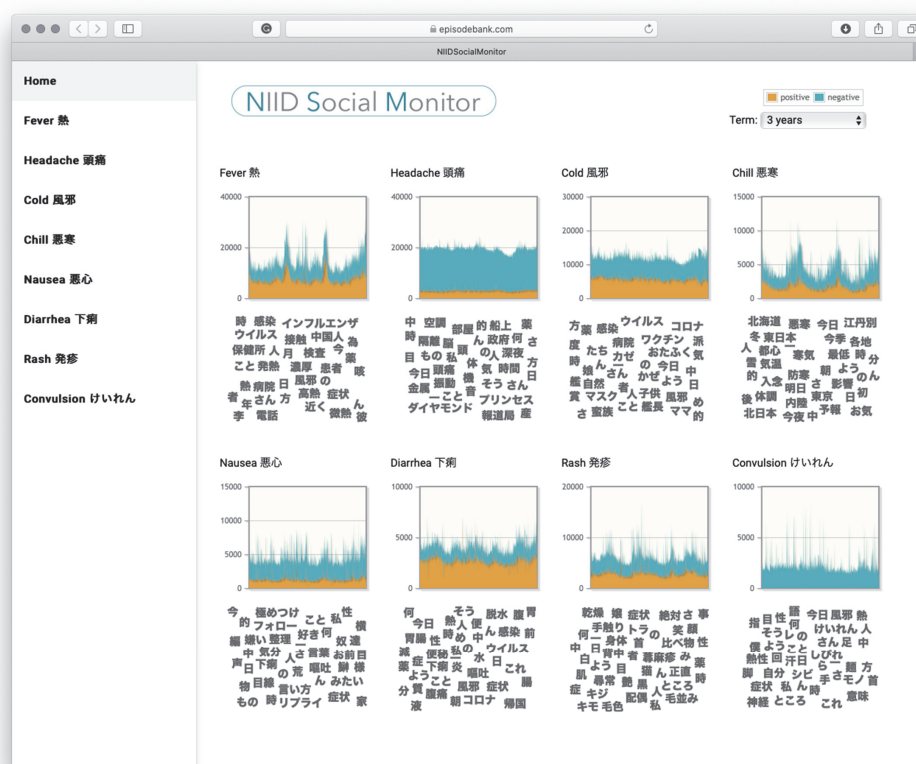


図 1. 症状ツイート可視化システムのユーザインターフェース。症状ごとに、ツイートの時系列を可視化したグラフ（横軸が時間、縦軸がツイート数、黄色の塗り潰しが陽性ツイート数、青色の塗り潰しが陰性ツイート数）とツイートから抽出された話題語を提示する。