

東京都内インフルエンザ流行状況の把握を目的としたツイート数の有用性の検討

北村 有里恵^a, 長島 真美^a, 吉田 勲^a, 根岸 あかね^a, 熊谷 遼太^a, 加来 英美子^a, 林 志直^a,
千葉 隆司^b, 秋場 哲哉^c, 新開 敬行^a, 貞升 健志^b

Social Networking Service (SNS) 上で発信される情報は多種多様な大量データ (ビッグデータ) により形成されており, 迅速かつ容易に収集可能なことから, 感染症分野での利用が期待されている. 本研究ではインフルエンザ等のウイルス性疾患に注目し, 流行状況の迅速な把握を目的としたTwitter上のツイート数の有用性の検討を行った.

2016年の「インフルエンザ」等のキーワードを含むツイート数と感染症発生動向調査におけるインフルエンザ患者報告数を比較したところ, ツイート数のピークを検知した時期は患者報告数のピークを検知した時期とほぼ一致したが, ヘルパンギーナ, 手足口病については, インフルエンザと比べ, 報道の影響が強く反映しており, ツイート数の増加は感染者の増加とは必ずしも連動しない結果となった. 感染症の流行把握には, ツイート数だけでなく, 感染症発生動向調査の患者報告数などの他のデータも併せて解析していく必要があり, 流行の早期探知についても引き続き解析を行っていく必要があると考えられた.

キーワード: インフルエンザ, サーベイランス, ビッグデータ, Twitter, リアルタイム, 早期探知

はじめに

Social Networking Service (SNS) 上でリアルタイムに発信される膨大な情報 (ビッグデータ) は, 迅速な情報収集が可能なことから, 商品開発などさまざまな分野で活用されている. 感染症分野においても荒牧らが, Twitter¹⁾ からインフルエンザに関連した発言を抽出し, Support Vector Machine (SVM) による文分類器を用いてインフルエンザの流行予測を行っている²⁾. また, 発言内容に含まれる位置情報を, 日本全国の地図にプロットすることでインフルエンザの流行地図を作成し, 「インフルくん」³⁾として情報を提供している.

2020年東京オリンピック・パラリンピックの開催など, 東京を中心に国内外からの大きな人の流れが予想されている. 人の移動に伴い, 新たな感染症の輸入, 感染拡大も想定され, 感染症発生の早期探知が極めて重要である.

今回, インフルエンザ流行の早期探知, 流行状況の把握を目的とし, リアルタイムに情報を得られ, コミュニケーションネットワークとして広く用いられているTwitterのツイート数を東京都におけるインフルエンザ患者報告数およびウイルス検出数の推移と比較検討した.

実験方法

Twitterに対象とするキーワードを含む投稿 (ツイート) がされた件数 (ツイート数) を日単位で集計した. 集計方法として, インターネットのYahoo! Japanの公式サイトから「リアルタイム検索」⁴⁾を用いた. 検索ワードを入力し, 検索結果のページより得られる分析グラフにてツイート数

のカウントを行った.

インフルエンザ関連キーワードとして「頭が痛い」「喉が痛い」「インフルエンザ」「インフルエンザA」「インフルエンザB」を用い, 2016年1月から5月を調査期間とした. また, 夏のウイルス性疾患であるヘルパンギーナ, 手足口病をとりあげ, キーワードは病名である「ヘルパンギーナ」「手足口病」を用い, 2016年4月から8月を調査期間とした.

インフルエンザ, ヘルパンギーナおよび手足口病の患者報告数は, 東京都感染症発生動向調査事業における患者定点医療機関から報告された患者数⁵⁾を用い, インフルエンザウイルス検出数は, 病原体定点医療機関から東京都健康安全研究センターに搬入された検体の遺伝子検索結果を用いた. ツイート数の推移と患者報告数およびウイルス検出数と比較し, 解析を行った.

結果と考察

1. 「頭が痛い」, 「喉が痛い」のツイート数の推移

「頭が痛い」および「喉が痛い」のツイート数の推移 (図1) をみると, 一日の平均ツイート数は「頭が痛い」が1,946件, 「喉が痛い」が1,326件で, 一過性の増減があるものの, 日常的にツイートされており, インフルエンザの流行を表す指標にはならないことが明らかだった.

2. 「インフルエンザ」ツイート数と患者報告数の推移

東京都ではインフルエンザ定点医療機関からの患者報告数が定点当たり1.0人を超えると「インフルエンザの流行

^a 東京都健康安全研究センター微生物部ウイルス研究科
169-0073 東京都新宿区百人町3-24-1

^b 東京都健康安全研究センター微生物部

^c 東京都健康安全研究センター企画調整部健康危機管理情報課

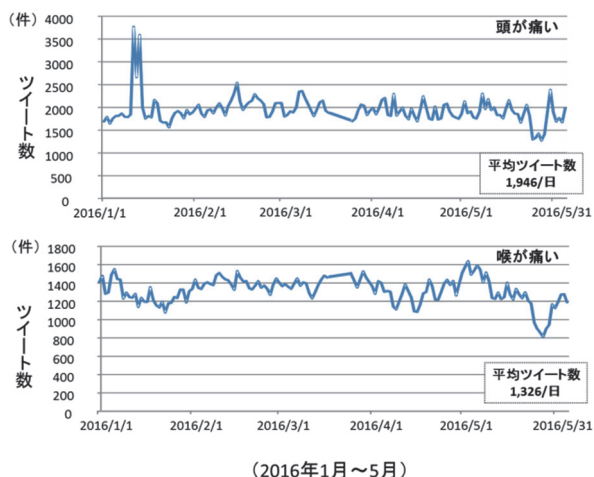


図1. インフルエンザ関連キーワード ツイート数の推移
「頭が痛い」「喉が痛い」

開始”，定点当たり10人を超えると“注意報”，定点当たり30人を超えると“流行警報”を発表している⁶⁾。2016年1月14日に“流行開始”，1月28日に“注意報”，2月12日に“警報”が発表され，それぞれの発表直後に「インフルエンザ」ツイート数は増加しており，報道がツイート数の増加に影響していることが明らかとなった（図2）。

「インフルエンザ」ツイート数とインフルエンザ患者報告数を比較したところ，ツイート数が最も多い週は第7週の176,182件，患者報告数のピークは第5週であり，ピークのずれがみられた（図2）。しかしながら，患者報告数は定点医療機関の受診日の翌週に集計して公表されるため，ツイート数のピークを検知した時期と患者報告数のピークが発表された時期はほぼ一致すると推察された。

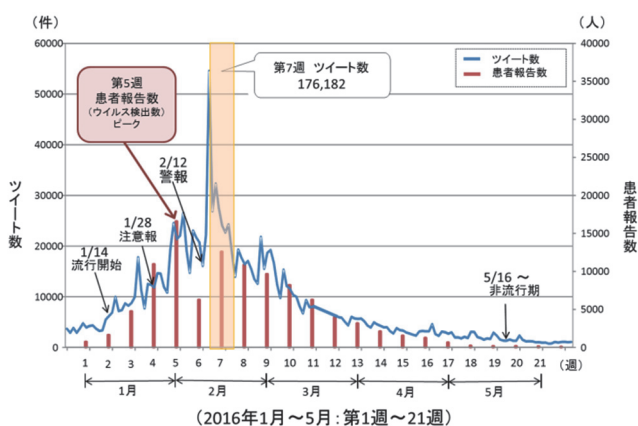


図2. 「インフルエンザ」ツイート数と患者報告数の推移

3. 「インフルエンザ A」，「インフルエンザ B」のツイート数，患者報告数とウイルス検出数の推移

「インフルエンザ A」および「インフルエンザ B」のツイート数と患者報告数の推移を示した（図3）。ツイート数の推移を見るときはじめは「インフルエンザ A」

が優位となるが，第7週に「インフルエンザ B」のツイート数が「インフルエンザ A」のツイート数を上回り，優位な型がA型からB型に変わっていった（図3）。このことから，第1～7週に報告された患者はA型が主流の流行であり，第7週以降はB型が主流の流行であることが予想された。

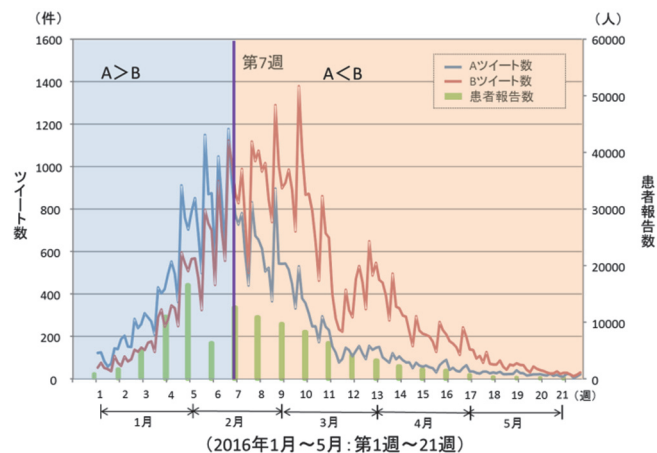


図3. 「インフルエンザ A」，「インフルエンザ B」のツイート数と患者報告数

次に「インフルエンザ A」および「インフルエンザ B」のツイート数とウイルス検出数の推移を示した（図4）。「インフルエンザ A」のツイート数は第6週に，検出数は第5週に最も多く，「インフルエンザ B」のツイート数は第9週に，検出数は第10週に最大となった。また，Aのツイートが優位な期間はA型のウイルスが優位に検出され，Bのツイート数が優位な期間はB型のウイルスが優位に検出されていたことが明らかとなった（図4）。

AとBの優位が切り替わる時期は，ツイート数では第7週，検出数では第6，7週であったことからインフルエンザのツイート数の推移と患者報告数，ウイルス検出数はほぼ一致した。

4. 「ヘルパンギーナ」，「手足口病」のツイート数と患者報告数の推移

インフルエンザ以外のウイルス感染症についてもツイート数と患者報告数の推移に関して比較検討を行った（図5）。「ヘルパンギーナ」では，患者報告数が23週から増えはじめ，30週をピークにその後減少していた。患者報告数が増えはじめた第24週からツイート数も増加した。第24週と第32週にツイート数の急激な増加がみられたが，これは第24週はヘルパンギーナの流行を伝えるニュースがネット上に流れたこと，第32週は厚生労働省がTwitterによりヘルパンギーナの注意喚起を行ったことに起因すると推測された（図5）。

「手足口病」では，患者報告数が少なく，警報レベルに当たる定点当たり5.0人を超えるような流行状況にな

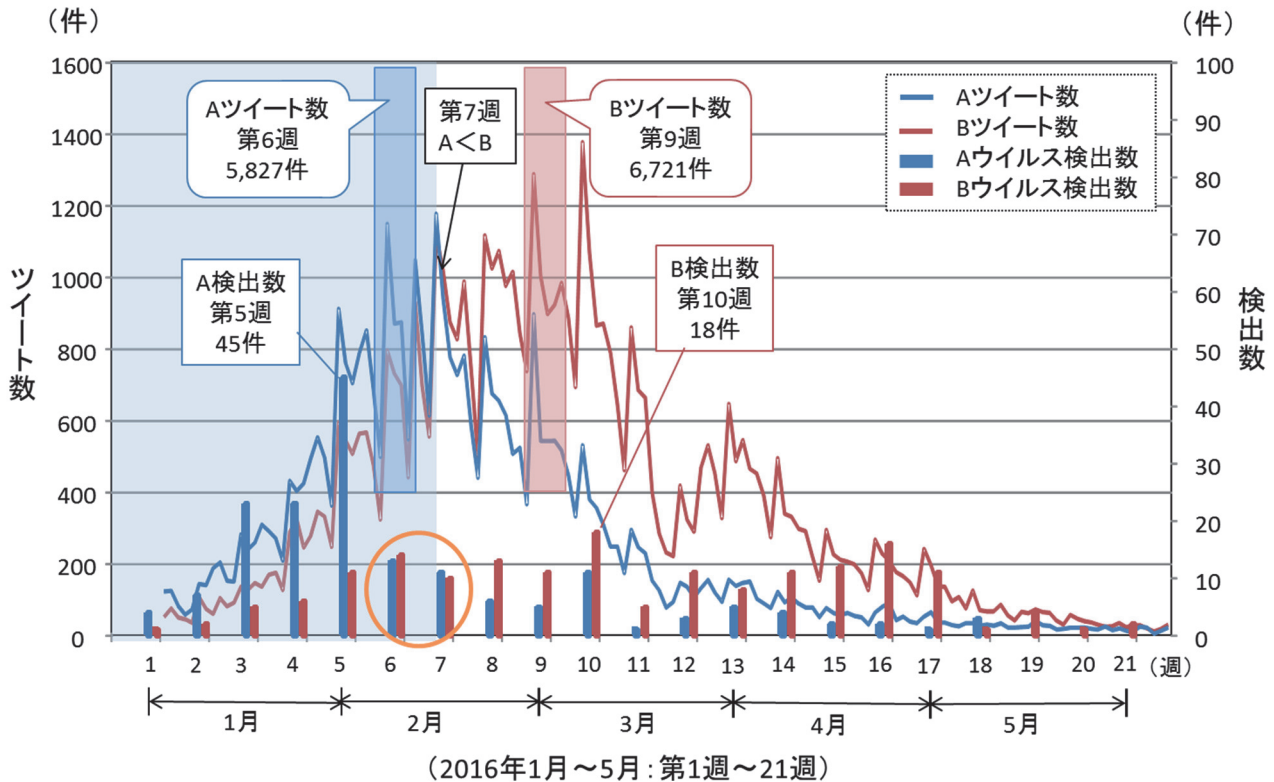


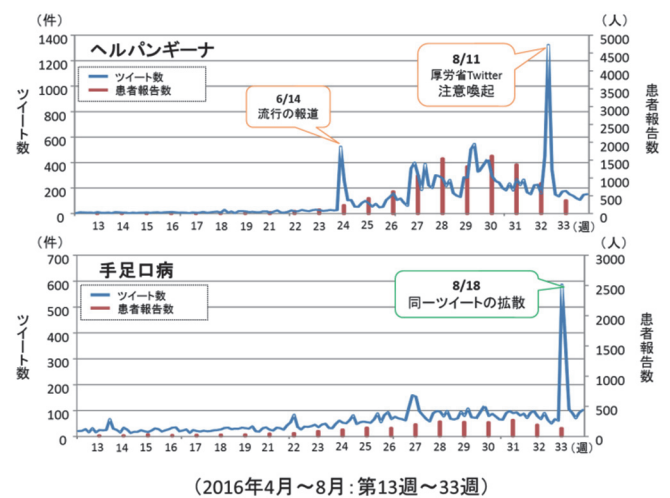
図4.「インフルエンザ A」,「インフルエンザ B」ツイート数とウイルス検出数

く、ツイート数も大きな変化はみられなかった。第33週にみられるツイート数の増加は手足口病に関する同一のツイートが多数拡散されていたことによる、一過性の増加によるものと推測された(図5)。

調査期間中、ヘルパンギーナや手足口病はインフルエンザに比べ大きな流行はみられなかった。ツイート数の急激な増加は、感染によるものではなく報道がきっかけになったものだったことから、ツイート数の増加が感染者の増加を必ずしも反映しないことが示唆された。また、ツイート数は全国でのツイートを対象としており、東京都の流行状況と一致しない場合も考えられる。石川⁷⁾は、患者の疫学的情報や発生数などを加えることにより、Twitterは強力なサーベイランスツールとなりうると考えられるとしている。今回の結果からも、ツイート数だけでなく、患者報告数などの流行状況を併せて解析していくことが重要と考えられた。

2020年には東京オリンピック・パラリンピックが開催され、国の内外から多くの人々が訪れるため、新たな感染症の発生が危惧される。和田らは海外からの訪日客が増加することで、患者数が増加する可能性がある疾患として、麻疹、風しん、インフルエンザなどを挙げている⁸⁾。

感染症発生時の早期探知は重要であり、早期探知を可能にする手法の一つとして、今後も継続して検討していく必要がある。

図5. ツイート数と患者報告数の比較
(ヘルパンギーナ、手足口病)

まとめ

Twitterに投稿されたインフルエンザ関連キーワードを含むツイート数の推移は、患者報告数、ウイルス検出数などの流行状況に関連していたが、すべての感染症に適用できるとはいえなかった。インフルエンザにおいて、A型、B型に分けて比較をしたところ、ツイート数とウイルス検出数の推移が一致しており、リアルタイムにツイート数の推移を見ることが、早期にインフルエンザ流行が探知できる可能性が示唆された。

文 献

- 1) Twitter社 : Twitter,
<http://twitter.com/> (2017年10月2日現在. なお本URL
は変更または抹消の可能性がある)
- 2) 荒牧英治, 増川佐知子, 森田瑞樹 : Twitter Catches
the Flu : 事実性判定を用いたインフルエンザ流行予測,
情報処理学会 第201回自然言語処理・第86回音声言語
処理 合同研究発表会 (東京), 2011年5月.
- 3) 荒牧英治 : 医療ビッグデータの深い言語解析による疾
患サーベイランスへ向けて, 第1回日本医療情報学会
「医用知能情報学研究会」・人工知能学会「医用人工
知能研究会 (SIG-AIMED)」合同研究会 (東京),
2015年9月.
- 4) Yahoo!Japan : リアルタイム検索,
<http://searchi.yahoo.co.jp /realtime> (2017年10月2日現在.
なお本URLは変更または抹消の可能性がある)
- 5) 東京都健康安全研究センター : 東京都感染症発生動向
調査,
<http:// survey.tokyo-eiken.go.jp/epidinfo/epimenu.do> (2017
年10月2日現在. なお本URLは変更または抹消の可能
性がある)
- 6) 東京都健康安全研究センター : インフルエンザの流行
状況 (東京都 2015-2016年シーズン),
<http:// idsc.tokyo-eiken.go.jp/diseases/flu/flu2015/>
(2017年10月2日現在. なお本URLは変更または抹消
の可能性がある)
- 7) Ishikawa, T.: *J. Natl.Public Health*, **61(5)**, 497-498, 2012.
- 8) 和田耕治, 西塚 至, 竹下 望, 他 : 東京2020オリ
ンピック・パラリンピック競技大会に關与する自治
体における感染症対策のためのリスク評価, 日医雜
誌, **145(7)**, 1459-1468, 2016.

Examination of the usefulness of the number of tweets for monitoring the influenza epidemic situation in Tokyo

Yurie KITAMURA^a, Mami NAGASHIMA^a, Isao YOSHIDA^a, Akane NEGISHI^a, Ryota KUMAGAI^a, Emiko KAKU^a, Yukinao HAYASHI^a, Takashi CHIBA^a, Tetsuya AKIBA^a, Takayuki SHINKAI^a and Kenji SADAMASU^a

Information transmitted via the Social Networking Service originates from a variety of large variety of data sources (big data) and can be collected in real time, therefore it is expected to be useful for the surveillance of infectious diseases . In this study, we focused on influenza and other infectious diseases and examined the usefulness of tweets on Twitter as an aid to monitor the epidemic situation.

We compared the number of “influenza” tweets in 2016 with the number of influenza patients reported in Tokyo. We found that the peak in the number of tweets almost coincided with the peak in the number of reported patients reports. However, with respect to, herpangina and hand, foot, and mouth disease, the increase in the number of tweets was attributed to news reports and was not necessarily linked with the increase in the number of infected individuals .

In order to detect outbreaks of infectious diseases, it is necessary to monitor not only the number of tweets but also the epidemic situation, such as the number of reported cases. Monitoring of tweets may help identify epidemiological trends and facilitate early detection of epidemics .

Keywords: influenza, surveillance, big data, Twitter, real time, early detection

^a Tokyo Metropolitan Institute of Public Health,
3-24-1, Hyakunin-cho, Shinjuku-ku, Tokyo 169-0073, Japan

