

# マイクロブログを用いた感染症サーベイランス

岡村直人<sup>†1</sup> 関和広<sup>†1</sup> 上原邦昭<sup>†1</sup>

感染症サーベイランスには、大別して医療機関の情報を利用した方法と Web 情報を利用した方法がある。Web 情報はリアルタイム性が高く、感染症の蔓延防止を目的とする感染症サーベイランスに有用である可能性がある。本論文では、Web 情報、特にソーシャルメディアを用いた感染症サーベイランスの第一歩として、インフルエンザを対象に Twitter の有効性を実験的に調査する。Twitter に投稿されたインフルエンザの症状を含む tweet の分類を行い、実際のインフルエンザ報告件数との関係を分析する。

## Using Microblog for Syndromic Surveillance

NAOTO OKAMURA,<sup>†1</sup> KAZUHIRO SEKI<sup>†1</sup>  
and KUNIAKI UEHARA<sup>†1</sup>

There are roughly two types of syndromic surveillance; One uses information from medical institutions and another gathers information from the Web. The information used by the latter, such as consumer generated media (CGM), may reflect more real-time events and thus may be more useful for syndromic surveillance since detecting early infection of a target syndrome prevents wider spread of the syndrome. This paper investigates the usefulness of CGM, specifically microblogs, for syndromic surveillance focusing on influenza. We collect a number of microblog posts (tweets) which include symptoms of influenza and study their size and transition against those of reported true influenza cases.

### 1. はじめに

近年、感染症の発生状況を調査・集計することにより、感染症の蔓延と予防に役立つ感

染症サーベイランスの研究が行われている。感染症サーベイランスには、大別して医療機関からの情報を用いたものと Web 情報を用いたものがある。Web 情報はリアルタイム性が高いため、一刻も早い感染症の発生検知を目的とする感染症サーベイランスに有用である可能性がある。Web 情報を用いた感染症サーベイランスの例として、Google Flu Trends<sup>1)</sup> がある。Google Flu Trends では、インフルエンザ関連の検索クエリを利用してインフルエンザの発生状況を推測する。本研究では、Twitter を用いた感染症サーベイランスの第一歩として、同様にインフルエンザを対象にその有用性を実験的に検討・調査する。なお、同様の試みとして tweetflu がある。tweetflu<sup>2)</sup> では、「インフルエンザ」という単語が含まれる tweet 数を都道府県別に集計し、各地域でどの程度インフルエンザの話題が上っているかを都道府県の地図上にマッピングし、可視化している。本研究は、注目する単語が「インフルエンザ」のみではない点、感染症サーベイランスにおける Twitter の有効性を定量的に調査している点などで、tweetflu とは異なる。

本論文の構成は次の通りである。2 章では、感染症サーベイランスと Twitter について説明し、Twitter を用いたサーベイランスに関する研究を紹介する。3 章では、本研究の方法の詳細について述べる。4 章では、本研究の評価として実データとの比較を行い、これについて考察する。最後に 5 章で、本論文のまとめについて述べる。

### 2. 関連研究

#### 2.1 感染症サーベイランス

感染症サーベイランスとは、「疾病の指標としての症状に着目し、その情報を自動収集し統計的手法により疫学的に解析することで、リアルタイムあるいはリアルタイムに近い状態で、疾病の発生をとらえる行為」のことである<sup>3)</sup>。感染症などの疾病をとらえることにより、感染症サーベイランスは感染症の蔓延と予防の支援を行うことを目的とする。しかしながら、医療機関からの情報を用いた感染症サーベイランスの場合、患者が症状を発症してから医療機関で診断を受けるまでにはタイムラグが生じる。そこで本研究では、比較的リアルタイム性が高いマイクロブログの情報を用いることで感染症発生の早期発見を試みる。次節では、本研究で利用したマイクロブログ Twitter について説明する。

#### 2.2 Twitter

Twitter は、現在最も盛んに利用されているマイクロブログサービスの 1 つである。Twitter を利用するユーザは、「今なにしてる?」という問いに 140 文字以内の短い記述で投稿（以下これを tweet と呼ぶ）することができる。tweet は、人々が何を見たか、聞いたか、思っ

<sup>†1</sup> 神戸大学大学院システム情報学研究所

Graduate School of System Informatics, Kobe University

たかなどを反映しているため、これらの情報を大規模に集約することにより実世界で起きている現象を捉えることが可能になると考えられる。

消費者が作成し発信する情報は一般に CGM (Consumer Generated Media) と呼ばれ、従来からブログや掲示板などの情報を対象とした研究やサービスが数多く存在する。ブログや掲示板と Twitter の大きな違いとして、以下の 2 点の特徴が挙げられる<sup>4)</sup>。

- リアルタイム性

ブログが平均的に 1 日に 1 回から数回しか投稿されないのに対し、Twitter は 1 日に数回～数 10 回投稿されることがしばしばあり、今何をしているか、今何を考えているかをリアルタイムに反映していることが多い。

- 利用場所

Twitter は、入力文字数が少ないことや今の状況を投稿するサービスの性質上、ブログと比較してモバイル端末から投稿されることが多い。

これらは、いずれも Twitter の特徴である 140 文字以内の短い記述で投稿できる気軽さに起因している。感染症サーベイランスにおいて、リアルタイム性の高い情報は重要であるため、従来の CGM よりも Twitter が感染症サーベイランスの情報源として適している可能性がある。

Twitter を対象とした研究も近年数多く行われている。例えば、Twitter についての調査<sup>5)</sup> やソーシャルネットワークの側面からの研究<sup>6)</sup>、コミュニケーションの種類<sup>7)</sup> などが挙げられる。Twitter を利用したサーベイランスという観点からは、前述の tweetflu の他にも花粉症観測システム<sup>4)</sup> がある。次節でこのシステムについて述べる。

### 2.3 Twitter を利用したサーベイランス

花粉症観測システムの目的は、花粉症の症状を訴える人々の地域毎の分布を天気予報の地図のように作成することである。システムは初めに「花粉症」というキーワードを含む tweet を取得し、そこから花粉症の症状を分類し、その結果を都道府県毎に集約し可視化する。可視化することにより、どの都道府県に花粉症の症状を含む tweet が多いかを容易に把握することが出来る。このシステムには、位置情報の推定や、数値の正規化、tweet の分類といったいくつかの技術的な課題がある。Twitter には、tweet に投稿した場所の緯度経度を付加できる tweet 位置情報という機能がある。ユーザは各 tweet 毎に位置情報を付加する/付加しないを選択することができる。この機能では、ユーザが tweet をする際に、近隣情報や町、または正確な現在地などの位置情報を含めることができる。しかし、全ての tweet に対して位置情報が付加されているわけではない。そこで、ユーザのプロフィールの

情報を用いて位置情報を取得する。Twitter ではユーザプロフィールの中に「位置情報」という項目があり、多くのユーザがここに居住地を書いている。

それぞれの都道府県には、人口の偏りに加え、Twitter を使っているユーザ数に偏りがある。また、一人当たりの tweet 数にも偏りがあると考えられる。都道府県毎の地図を描くためには、この分布の異なりを吸収するために tweet 数の正規化を行わなければならない。この正規化のために、Twitter ユーザの都道府県別の分布を調査する。ここで用いる都道府県名は、tweet に緯度経度情報が付与されていれば、位置情報に対応する都道府県名を取得し、付与されていなければ、前述で説明したプロフィールの位置情報から都道府県名を推定する。取得した tweet 数は全都道府県での最大の tweet 数で比をとる。例えば、tweet 数の重みが 100 であればある県が他の県に対して tweet 数が 100 倍であることを表している。

このシステムでは、まず tweet を「花粉症」というキーワードにより収集する。しかし、これらの tweet を投稿したユーザのすべてに花粉症の症状があるとは限らず、「花粉症」というキーワードが含まれているだけの場合もある。例えば、「花粉症にはなりたくない」や「花粉症が流行ってるので注意したほうがいいよ」などの tweet である。これらの tweet を区別するために、花粉症の症状がある/ない、という 2 つのカテゴリに tweet を分類し、症状があるとされた tweet のみを用いている。しかし、花粉症であることと、花粉症の症状があることは表層的な情報だけでは区別が難しく、分類の精度も 77.27% と (二値分類としては) それほど高くない。

## 3. 研究の方法

### 3.1 概要

本研究の目的は、感染症サーベイランスにおける Twitter の有効性を調査することである。そのために、インフルエンザの症状に関連する tweet を都道府県別に調査し、実際のインフルエンザ報告件数と比較する。まず、クエリを含む tweet を場所別に取得し、正規化を行う。その後、tweet をインフルエンザの症状があるかないかで分類する。tweet 取得のための API の仕様やクエリ選別、場所の決定といった課題については、以降で述べる。

### 3.2 tweet の取得

本研究では、tweet 数の場所毎の異なりについて調査するため、場所を指定して tweet を取得する必要がある。そこで、Twitter 社が提供する API <sup>\*1</sup> を利用して、場所を明示的に

\*1 <http://apiwiki.twitter.com/>

指定して tweet を取得する。本研究では、特に日本の地域を場所指定して tweet が取得できる Search API を用いる。tweet の取得には他に Streaming API があり、これを利用すれば指定した場所で投稿された tweet を取得することができる。しかしながら、この API は日本の地域には対応していないため、本研究では使用しない。Search API のオプションのうち、本研究の tweet の取得に必要なものを以下に挙げる。

- q  
クエリを指定する検索条件である。クエリが含まれる tweet を対象にする。クエリが指定されていない場合は、全ての tweet が対象になる。
- geocode  
緯度と経度と半径を指定する。その緯度と経度を中心地として、指定した半径内で投稿された tweet を対象にする。
- since\_id  
指定した tweet ID よりも後に投稿された tweet を対象にする。

主にこの 3 つのオプションをそれぞれ決めることで、tweet を取得する。

### 3.2.1 API の問題点

API には仕様上の問題がいくつかある。以下に本研究に関連した問題点を挙げる。

- API 制限  
Twitter 側が負荷軽減のために設けているデータ取得系 API の 1 時間当たりの取得回数制限である。1 時間に 150 回を超える API の利用はできない。
- 取得 tweet 数制限  
最大で取得できる tweet 数は 1500 件となる。tweet は投稿時間が新しいものからしか取得できず、1500 件より前に投稿された tweet は取得することができない。
- クエリ数制限  
Search API ではリクエストの文字数が 140 文字と制限されている。Twitter API では UTF-8 エンコーディングを使用するため、クエリがエンコードされた状態で文字数をカウントする。一般的にエンコードされた日本語は、元の文字数よりも多くなるためクエリとして利用できる文字数が少なくなる。

### 3.2.2 対象となる tweet の決定

tweet を取得する際に、前述のクエリ、位置、since\_id の 3 つを決定する。クエリに tweet であまり使われない表現を用いると、該当する tweet はほとんどない。そのため、まずクエリ候補を作り、その中から tweet に含まれる割合の多い単語だけを使用する。クエリの候補

は、厚生労働省が発表している新型インフルエンザに係る症例定義<sup>8)</sup> の中に含まれる単語とする。また Twitter では、文語的な表現が少なく口語的な表現が多いため、上記の単語を話し言葉に言い換えた表現も候補とする。例えば、症例定義に含まれる「倦怠感」は、「だるい」という表現で使われるため、「全身倦怠感」と「だるい」の両方を候補する。クエリを決定するため、2010 年 12 月 18 日～2011 年 1 月 24 日（計 38 日間）に取得した 716,417 件の tweet 中に各クエリ候補が含まれていた件数を表 1 に示す。全体の tweet 数が 70 万に

表 1 各クエリ候補と tweet に含まれている件数。

クエリ候補	件数
熱	2849
風邪	2486
インフル	651
頭痛	642
咳	393
インフルエンザ	336
筋肉痛	284
だるい	269
鼻水	176
かぜ	119
下痢	102
くしゃみ	98
寒気	97
嘔吐	58
悪寒	55
高熱	29
38	8
倦怠感	6

も及ぶため、件数が 2 桁以下の単語はほとんど tweet に含まれていないと考えられる。クエリ文字数の制限から、表 1 の候補の中から件数が多い上位 5 個のクエリ候補を使用する。選ばれたクエリは、「咳」「頭痛」「インフル」「風邪」「熱」の 5 個である。

次に位置の決定を行う。その際には API 制限を考えなければならない。API を 1 回利用することで得られる最大の tweet 数は、Search API では 100 件となっている。よって tweet 取得最大件数の 1500 件の tweet を得るためには API を 15 回利用しなければならず、1 箇所につき API の利用回数が 15 回必要である。都道府県の全ての場所を対象にすると tweet

の取得回数で API の制限を超えるため、tweet を取得する場所を絞り込む必要がある。そこで場所の決定のために、日本医師会が公開している感染症サーベイランス<sup>9)</sup> のデータを用いる。この感染症サーベイランスは、全国の参加医療機関から報告された過去 3 日分のインフルエンザ診断件数を地図上に表示するシステムである。このサーベイランスシステムを用いて各都道府県での毎日のインフルエンザ診断件数を調べ、インフルエンザ診断件数が特徴的な場所（多い場所と少ない場所）を探す。本研究では、インフルエンザ報告件数の多い 4 箇所と少ない 3 箇所を tweet の取得場所とする。表 2 に tweet を取得する都道府県と 3 日間のインフルエンザ報告件数の集計を示す。インフルエンザ報告件数の多い都道府

表 2 tweet を取得する都道府県の日ご別インフルエンザ報告件数。

都道府県	インフルエンザ報告件数	
	1月31日~2月2日	2月3日~2月5日
東京	211	123
大阪	271	95
岡山	560	281
福岡	258	224
岩手	21	6
三重	6	10
香川	14	7

県では、3 日間集計で少なくとも 100 件程度あり、少ない都道府県では多くて 20 件程度ある。tweet を取得する緯度経度は、それぞれの都道府県の県庁所在地を中心地に設定する。緯度経度を指定して、中心地から指定した半径内に含まれる tweet を対象にすることで、指定した位置周辺の正確な tweet を取得することができる。tweet の取得半径を一律に決めてしまうと各都道府県で tweet 数の差が大きくなる可能性がある。この問題に対処するため、API を利用する際にクエリを指定せずに、それぞれの都道府県で 1 件目の tweet と 1500 件目の tweet の投稿時間の差が 24 時間程度になるように半径を決定する（以下、クエリを指定せずに取得された tweet 件数を母数と呼ぶ）。

半径を上記のように 24 時間で 1500 件程度になるように決定しているため、1 日毎に tweet を取得する。なお、tweet は投稿の新しいものから取得され、当日の tweet が 1500 件に満たない場合、1 日前やそれ以前の tweet も取得する。よって、以前に取得された tweet を再取得してしまう可能性がある。本研究では、インフルエンザに関連した tweet の数を日に

ち別に調査しており、tweet が重複してしまうと正確な数が分からなくなる。そこで tweet ID を利用する。tweet ID は、tweet に付与された一意の番号であり、投稿が新しい tweet ほど ID は大きくなる。そのため、前日に取得した tweet で最も投稿時間が新しい tweet の ID を調べ、その ID を since\_id とする。これにより、取得される tweet は前日に取得した tweet 以降に投稿されたものになるので、tweet の重複を避けることができる。

### 3.3 正規化

tweet を取得する都道府県は、前節のように範囲を決定しているため、日によってそれぞれ母数が若干異なる。母数が異なると、インフルエンザに関する tweet 数が同じでも 1 件に対しての重みが変わってくる。例えば、母数が 1500 件でインフルエンザに関連する tweet 数が 100 件取得できた場所と、母数が 500 件でインフルエンザに関連する tweet 数が 100 件取得できた場所では、インフルエンザに関連する tweet の割合が異なる。そこで、この異なりを吸収するために tweet 数を正規化する。具体的には、各日のそれぞれの都道府県でのクエリを指定しない時の tweet 数も調べ、この数と基準値（1500 件）との比を求める。この比をインフルエンザに関連した tweet 数に加味する。

### 3.4 tweet の分類

本研究では、まず「咳」「頭痛」「インフル」「風邪」「熱」の 5 個のクエリを用いてインフルエンザに関する tweet を収集する。これにより、5 個のクエリいずれかの表現を含む tweet が取得できる。しかし、これらの tweet を投稿した全てのユーザにクエリに対応する症状があるわけではない。例えば、「最近インフル流行ってるから注意してねー。」や「風邪になるのは嫌だ!。」といった単にクエリが含まれている tweet がその例である。また、「息子がインフルにかかってしまった」などの投稿者自身は症状がないものの、家族には症状があるなどといった tweet もある。これらの tweet を区別するために、クエリの症状がある/ない、という 2 つのカテゴリに tweet を分類し、症状があるとされた tweet だけをそのクエリの tweet の数に含める。

本研究では、クエリの表現を含む症状の有無を表 3 の 4 つのパターンに分類する。このパターンの中から 1, 2 を症状あり（陽性）、3, 4 を症状なし（陰性）として分類する。2 も陽性に含めた理由は、投稿者の家族は同じ場所に住んでいることが多いと考えられるためである。本研究では、都道府県毎の tweet 数の異なりを調査しているため、投稿者の家族であっても同じ場所に住んでいれば tweet 数にカウントすることは妥当である。なお、3 の場合は、他人が同じ都道府県に住んでいるかどうかは分からない。例をあげると、「そっちはインフルが〜。大変だね。」という tweet が東京で投稿されたとすると、症状がある人は沖

表 3 クエリの表現を含む症状のパターン.

パターン	説明
1	投稿者自身が症状あり
2	投稿者の家族が症状あり
3	他人が症状あり
4	症状なし

縄に住んでいるかもしれないし、北海道に住んでいるかもしれない。このような tweet を陽性に含めると、投稿場所と異なる場所も含めた tweet 数になってしまうおそれがある。

#### 4. 評価

本章では、3章で述べた方法の評価として、各クエリの妥当性評価と実際のインフルエンザ報告件数との比較を行う。また、各都道府県でインフルエンザ報告件数と tweet 数の関連を調べる。

##### 4.1 クエリの妥当性評価

各クエリを含む tweet が、実際にどれほどインフルエンザの症状あり（陽性）の場合に含まれるのかを調べた。陽性である tweet の割合が低ければ、そのクエリは有効ではないと考えられる。これを調べるために、各クエリを含む 1500 件の tweet をそれぞれ用意した。1500 件の中から 200 件の tweet をランダムサンプリングし、人手によって症状のあり/なしの分類を行った。表 4 に各クエリの分類結果を示す。クエリの中で関連が高かったのは

表 4 各クエリに関するインフルエンザの症状に関連する割合.

クエリ	症状ありの件数	症状の関連する割合
頭痛	129	0.645
咳	117	0.585
風邪	62	0.310
熱	48	0.240
インフル	39	0.195

「咳」と「頭痛」の 2 つである。この 2 つのクエリの tweet の例は、「まだ咳が止まらない。明日も止まらなかったら病院に行こう..」、「頭痛う。しんどいよ 精神的に 昨日話きいてもらったのに。どんどん落ちていくのがわかる」などである。このような tweet から「咳」

と「頭痛」というクエリは、自分の症状を説明している tweet が多いということが分かる。また、関連が低かった「インフル」「風邪」「熱」について、どのような tweet が多かったのかを調べた。その結果、「インフル」や「風邪」というクエリには、「@mhadate 風邪? だいいょぶー??」、「最近寒いけど風邪とか引くんじゃねーぞ」、「汗は拭かなあかんよ! インフルも流行中」などの tweet が多かった。これらにより、相手を気遣う tweet が多く、自分の症状についての tweet が少ないことが関連が低い原因だと考えられる。また、「熱」というクエリについては、「餃子が超熱い!」、「金曜日まで塾あるとか胸熱」、「あああ友人のおかげでまたハガレン熱がきた。シン国組が異様好きだ」などの症状に全く関係ない tweet が多かった。これらから、「インフル」や「風邪」のような病名よりも、「咳」や「頭痛」といった症状をクエリにしたほうが、投稿者自身の症状がある tweet を抽出できるものと考えられる。

また、今回用意した 1500 件の tweet の中には、@で始まる reply（ある tweet に対する返事）、RT で始まる retweet（他の Twitter ユーザの発言を自分のアカウントで再投稿すること）が含まれていた。reply や retweet は、通常の tweet と比べて他人に関する内容の tweet が多く、自身の症状についての tweet が少ない。ランダムサンプリングした 200 件の tweet の中にもこの reply と retweet が多数含まれていたため、これが全体的に症状の関連する割合が低い原因だと考えられる。クエリに対する症状の関連する割合を高くするためには、reply や retweet を取得せずに、通常の tweet のみを取得することが必要である。

##### 4.2 各地域での比較

インフルエンザ報告件数の多い都道府県と少ない都道府県でそれぞれ得られたインフルエンザの症状が含まれる tweet 数の比較を行った。インフルエンザの症状が含まれる tweet 数を報告件数が多い都道府県と少ない都道府県で図示した。図 1 に報告件数が多い都道府県の tweet 数の推移を示す。次に、図 2 に報告件数が少ない都道府県の tweet 数の推移を示す。図 1 と図 2 の 2 つから、インフルエンザ報告件数が多い都道府県は、インフルエンザ報告件数が少ない都道府県に対してインフルエンザの症状に関連する tweet 数が若干多いということが分かる。これにより、tweet 数が多ければその都道府県でのインフルエンザ感染者数が多い可能性が高いと言える。

しかしながら、今回取得された tweet 数は少なく、十分な比較が行えなかった。この原因は、3.2.2 節で述べた範囲を決め方だと考えられる。母数が 1500 であるため、この中からクエリを含む tweet 数はさらに少なく、今回の結果のように多くても 20 程度しかない。よって範囲を広げて、取得できる tweet を増やすなどの対応が考えられる。この際には、隣り合った県を範囲の中に含まないようにしなければならない。また、緯度経度を指定した位置

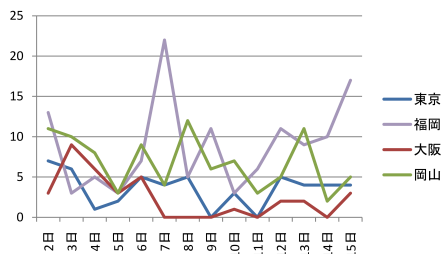


図1 報告件数が多い都道府県の tweet 数の推移.

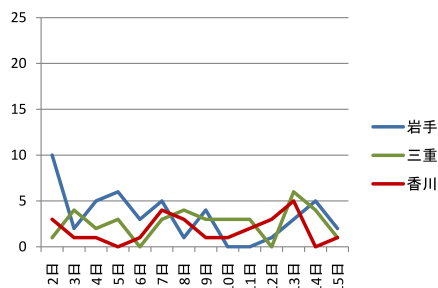


図2 報告件数が少ない都道府県の tweet 数の推移.

は各都道府県の県庁所在地になっており、それぞれの中心地ではない。そのため、半径を指定してその半径に含まれる tweet を対象にする本研究では、各都道府県で範囲に含まれていない地域が存在する。その地域で Twitter を利用している人口が多ければ、これを対象としないことで tweet 数が正確でないということが言える。実際のインフルエンザ報告件数のデータでは、それぞれの都道府県の地域全体の医療機関から報告されているため、取得の対象にした tweet が含まれる地域が異なる。

#### 4.3 インフルエンザ報告件数との比較

tweet から得られたインフルエンザの情報の評価のために、実際のインフルエンザ報告件数との間に関係があるかどうかを分析した。まず、実際のインフルエンザ報告件数のデータとして、3.2.2 節で述べた感染症サーベイランスを用いた。このシステムでは、各都道府県の日毎のインフルエンザ報告件数があり、本研究で指定した都道府県に対してのデータを用いた。

3 節で述べた方法により取得した、2011 年 2 月 2 日～2011 年 2 月 15 日までの計 14 日間におけるインフルエンザの症状が含まれる tweet と実際のインフルエンザ報告件数との相関係数を都道府県毎に分析した。この tweet 数は 3.3 節で述べた正規化を行った後、小数点以下を四捨五入した値である。感染症サーベイランスという観点からは、報告件数と同日の tweet 数よりも 1 日前、2 日目の tweet 数との相関があるほうが、それだけ感染を早期に検出していることになるため好ましい。そこで、報告件数と 1 日前、2 日前、3 日目の tweet 数の相関係数を求めた。図 3 に、報告件数と tweet 数との都道府県毎の相関係数の推移を示す。この結果から、報告件数と同日の tweet 数より 1 日目の tweet 数との相関が強いこ

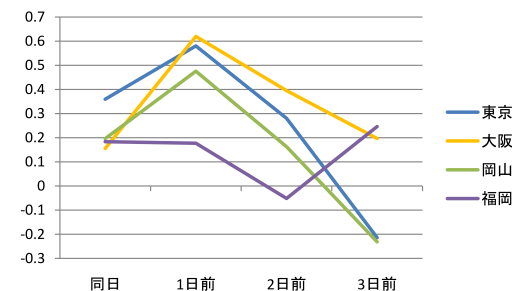


図3 報告件数と tweet 数との都道府県毎の相関係数の推移.

とが分かる。この結果は、Twitter を用いることで、医療機関に患者が来院する以前にインフルエンザの感染拡大が検知できる可能性を示している。

#### 5. 結 論

本論文では、Twitter を利用した感染症サーベイランスについて、インフルエンザに注目して調査・検討を行った。インフルエンザの症状をクエリとして、クエリが含まれる tweet を都道府県毎に取得し、症状のあり/なしでの tweet の分類を行った。また、実際のインフルエンザ報告件数と本研究の方法で取得した tweet 数との比較を行った。その結果、報告件数と同日の tweet 数より 1 日目の tweet 数との相関が強いことが分かった。この結果は、Twitter の利用によってインフルエンザの感染拡大が早期に検知できる可能性を示すものである。今後は、本研究の方法で述べた取得の対象となる tweet が含まれる範囲を大きくし、さらに大規模な比較を行う必要がある。また、クエリを含む tweet に関して、症状のあり/

なしの分類を各クエリで行った結果，病名をクエリにするよりも症状をクエリにした方が効果的であることが分かった．感染症発生の検知精度をさらに高めるためには，質の低いクエリを除外，あるいは重み付けして利用するなどの方策が考えられる．

### 参 考 文 献

- 1) Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. and Brilliant, L.: Detecting influenza epidemics using search engine query data, *Nature*, Vol.457, pp.1012–1014 (2009).
- 2) 西藤なるを：tweetflu, available from <http://tweetflu.jp/> (accessed 2011-02-12).
- 3) 重茂浩美：症候群サーベイランス-感染症流行の早期探知に向けて-，科学技術動向，No.109, pp.8–22 (2010).
- 4) 高橋哲郎，野田雄也：実世界のセンサーとしての Twitter の可能性，信学技報 NLC2010-38，Vol.110, No.400, pp.43–48 (2011).
- 5) Java, A., X.Song, T.F. and Tseng, B.: Why we twitter: understanding microblogging usage and communities, *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, (WebKDD/SNA-KDD 2007)*, pp.56–65 (2007).
- 6) Kwak, H., Lee, C., Park, H. and Moon, S.: Whata is twitter, a social network or a news media?, *Proceedings of the 19th International Conference on World Wide Web, (WWW 2010)*, pp.591–600 (2010).
- 7) Naaman, M., Boase, J. and Lai, C.-H.: Is it really about me?: message content in social awareness streams, *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW 2010)*, pp.189–192 (2010).
- 8) 厚生労働省：新型インフルエンザに係る症例定義，  
入手先(<http://www.mhlw.go.jp/kinkyu/kenkou/influenza/090429-03.html>)  
(参照 2010-12-05)
- 9) 日本医師会：感染症サーベイランス (インフルエンザ)，  
入手先([http://www.orca.med.or.jp/das/infection\\_map/old\\_maps.html](http://www.orca.med.or.jp/das/infection_map/old_maps.html))  
(参照 2010-10-28)