
一般口演

一般口演5 知識工学

2017年11月21日(火) 08:30 ~ 10:00 G会場 (10F 会議室1006-1007)

[2-G-1-OP5-5] 症状に関するつぶやき数の多重線形回帰による感染性胃腸炎 流行規模の推定

竹内 瞭, 伊藤 薫, 若宮 翔子, 荒牧 英治（奈良先端科学技術大学院大学）

【背景・目的】感染症は例年、百万人を超える患者の規模で流行しており、流行の抑制が重要な課題となっている。流行の抑制を実現するためには、感染症の流行規模や患者数推移の早期把握などが課題に挙げられ、対策として感染症サーベイランスと呼ばれる感染者数の集計が実施されている。しかし、感染症サーベイランスは連携医療施設からの報告に依存しているため、集計に時間が必要である。そこで近年、SNS (Social networking service) を用いた感染症サーベイランスが注目されている。SNSを通して、インフルエンザなど大規模かつ季節的な感染症の流行を早期把握する試みは一定の成果が報告されている。一方で、小規模かつ突発的な活動に対しては十分な調査がされておらず、検討の余地がある。そこで本研究では、日本において二番目に報告数が多く、突発的に流行する感染性胃腸炎を対象として、当該疾患の流行を SNSによりどの程度の推定が可能かの検討を行う。

【手法】日本国内で SNSにおける感染性胃腸炎の症状に関するつぶやきを、機械学習を用いて陽性、陰性の二値分類を行う。次に陽性と判断された発言に対して多重線形回帰を適用し、推定を行った。

【結果・考察】線形回帰のデータには、2016年の3月～12月の間に群馬県でつぶやかれたものを用いた。2017年1月～4月における各都道府県の週毎の罹患率と推定結果を4週間毎に区切り、2つのデータの相関係数を求めた結果、患者数の上昇規模と推定精度との相関は0.13にとどまった。しかし、感染者数が上昇傾向にある区間かつ人口が200万程度の都道府県に限定した場合、患者数の上昇規模と推定精度との間に一定の相関がみられた（相関係数0.38）。このことから、人口200万程度の都道府県において1カ月で150人以上の患者の増加が見られるようなケースでは、強い確信を持って患者数の上昇規模の推定が可能であることが示された。

症状に関するツイート数の多重線形回帰による感染性胃腸炎流行規模の推定

竹内 瞭^{*1}、伊藤 薫^{*1}、若宮 翔子^{*1}、荒牧 英治^{*1}

*1 奈良先端科学技術大学院大学

Twitter-based Epidemics Detection

Ryo Takeuchi^{*1}, Kaoru Ito^{*1}, Shoko Wakamiya^{*1}, Eiji Aramaki^{*1}

*1 Nara Institute of Science and Technology

To-date various Twitter-based event detection systems have been proposed. Most of their targets, however, share common characteristics. They are seasonal or global events such as earthquakes and flu pandemics. In contrast, this study targets unseasonal and local disease events. Our system investigates the frequencies of disease-related words such as “nausea,” “chill,” and “diarrhea” and estimates the number of patients using regression of these word frequencies. Experiments conducted using Japanese 47 areas from January 2017 to April 2017 revealed that the detection of small and unseasonal event is extremely difficult (overall performance: 0.13). However, we found that the event scale and the detection performance show high correlation in the specified cases (in the phase of patient increasing or decreasing). The results also suggest that when 150 and more patients appear in a high population area, we can expect that our social sensors detect this outbreak. Based on these results, we can infer that social sensors can reliably detect unseasonal and local disease events under certain conditions, just as they can for seasonal or global events.

Keywords: Twitter, Event detection, Gastroenteritis, Social media, Natural Language Processing

1 はじめに

インフルエンザや感染性胃腸炎などの感染症は、予防のために早期把握が課題となっている。このような背景中、情報処理分野においても、近年普及している Twitter などのソーシャルメディアをソーシャルセンサ¹⁾とみなし、感染症の流行把握に活用する研究が注目を集めている。従来の研究は、インフルエンザ^{1), 12), 16), 19)} のように、1 年間に 1 回時には数万人規模で流行するような季節性疾患を対象としたものが多い。このような大規模な流行はソーシャルメディア上での患者による発言量が増加する傾向があるため把握は容易である。例えば、ソーシャルメディアを代表する Twitter において、インフルエンザに関するツイートは一日当たり 10 万件を超えることがある。周期性がなくとも、大規模に発生する大腸菌⁵⁾ や H1N1-type flu^{4), 11)} などについても同様である。

一方、小規模かつ季節性のない流行を引き起こす感染症も存在する。このタイプの感染症の主な例として、感染性胃腸炎がある。図 1 にインフルエンザと感染性胃腸炎の患者数に基づく流行度を比較したものを示す。多くの場合、感染性胃腸炎はノロウイルスやカンピロバクターなどのウイルスによって引き起こされ、下痢、嘔吐、発熱、脱水症状などの症状が通常 2 週間程度続く。インフルエンザとは異なり、季節性はなく、どの季節でも流行(以降、アウトブレイクと呼ぶ)する可能性がある。さらに、アウトブレイクは1回きりというわけではなく、複数回流行する可能性がある。感染性胃腸炎の流行規模は、数人規模の小さなものから数千人規模の大きなものまで多様である。実際の病名を用いてつぶやくこと、例えば「感染性胃腸炎なう」、はほとんど見られない。インフルエンザと感染性胃腸炎の流行の比較を図 1 に示す。

1. 非季節的流行: 感染性胃腸炎の発生には季節性はなく、どの季節でもアウトブレイクする可能性がある。さらに、1 年間に多くのピークが存在する可能性がある。

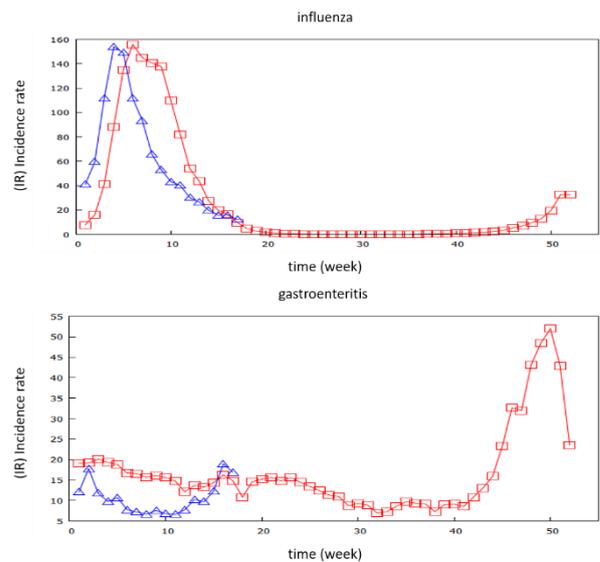


図 1. 全国規模かつ季節的流行(a) vs 地域規模かつ非季節的流行 (b). X 軸は時間列(週毎), Y 軸は罹患率(IR)を示しており、日本国内における 2016 年と 2017 年の疾患の流行推移を示している。赤線は 2016 年、青線は 2017 年の流行推移をそれぞれ示している。(a) 全国規模かつ季節的流行(日本国内でのインフルエンザ)。インフルエンザは 1 年を通して流行のピークは 1 回のみであることがわかる。(b) 地域規模かつ非季節的流行(感染性胃腸炎)。感染性胃腸炎は 1 年を通して不定期に複数回の小規模な流行があり、予測することは難しい。

2. 地域規模での流行: 感染性胃腸炎の流行規模は、数人規模の小さいものから数千人規模の大きなものまでさまざまである。

これらの特徴より、インフルエンザの流行検出のための方法をそのまま感染性胃腸炎の流行に適用することは困難である。

そこで、本研究では、比較的小さい規模の流行での推定性能とその性能に影響する要因を調査した。この結果、流行規模がソーシャルセンサの推定精度に影響を与える主な要因であることが示され、人口密集地において4週間に患者数が約150人程度増加した際、高い推定精度が検出された(ソーシャルセンサの患者推定数と実際の患者数との相関は0.8であった)。

この結果から小規模な事象の検出は困難であるが、もし、150人以上の大規模な流行が発生した際には高い確率で検出が可能であり、これは、実用上、大きな利点となると考えている。

2 関連研究

感染症の早期把握は、国家の重要な課題の1つである。このとき、早期把握すべき感染症は大きく次の2つに分類される。(1) インフルエンザなどの季節性感染症。(2) 食中毒(感染性胃腸炎)やバイオテロなどの非季節性感染症。感染症の早期把握のために、多くの国ではそれぞれの感染症予防センターが存在し、米国では感染症対策予防センター(CDC), E.U. では欧州インフルエンザサーベイランススキーム(EISS), 日本では感染症サーベイランスセンター(IDSC)が存在する。それぞれの感染症予防センターはウイルス学と臨床データに依存している。例としてIDSCは5,00の診療所からインフルエンザの患者データを収集し、報告を公表している。このようなシステムは、通常、1~2週間の報告の遅延が発生し、重大な欠陥としてしばしば指摘されている。感染症の早期把握する試みとして今まで様々な手法が提案されている(6), 14)。

Webベースを用いた最初の感染症サーベイランスはGoogle Flu Trends(GFT)である。これはGoogleのクエリログのデータセットを用いて、インフルエンザの患者数を予測する手法である(7)。GFTはWebベースによるサーベイランスの有効性を示したが、Googleのクエリログは公開されたデータセットではないため、一般で用いることはできない。

最新のWebベースによる感染症サーベイランスシステムは、主にTwitterなどの公開データセットに依存している(8), (13), (16), (18), (19)ではいくつかのインジケータからの情報リソースを用いて、米国、イタリア、スペインで得られた予測精度を報告している。日本で最初の成功例は(1)である。これはインフルエンザ関連の単語を含むツイートごとに、インフルエンザに感染しているかどうかの分類を行っている。(18)は、さらに細かい感染症の把握のために都市と地方都市での人口の違いについて検証を行っている。日本の感染症モデルを用いた最新の研究である(8)は、予測モデルの改善のためタイムラグを用いた。しかし、これは日本全体の患者数を調べるだけで、地域ごとの人口の規模を考慮していない。本稿では、Twitterベースの推定精度の調査のため、数人規模から数千人規模の感染性胃腸炎流行まで、様々な規模の流行を対象にして検証を行った。

3. 手法

3.1 ツイートの収集と分類

ツイートから感染性胃腸炎の流行を検出するために、最初に以下の3つのステップから患者数の推定を行った。

はじめに、Twitter API からツイートの収集を行った。このとき、感染性胃腸炎の典型的な症状である“嘔吐”、“悪心”、“下痢”の3症状をキーワードとして設定した。このキーワードセットは既存研究の実験(2)で用いられた11の症状の中から選択した。

次に収集したツイートコーパスを用いて、特定のツイートが患者によってつぶやかれたものかどうか(陽性か陰性)を判定する分類器を構築した。本研究では、bag-of-words(BOW)から、SVMを用いて分類を行った。このとき、我々はツイートをMeCab2(ver.0.98)とIPADic(ver.2.7.0)¹⁰⁾を用いて、単語ごとに分割を行った。このとき、多項式カーネル($d=2$)がカーネル関数として扱われる。その他のパラメータは、以前の報告書のものを用いる。トレーニングセットを構築する際、人手で陽性または陰性のラベルを割り当てる。ラベリングの際、以下の2つの条件を満たすツイートを陽性と判断した。表1にラベルを付与したツイートの例を示す。

条件1: 患者か否か

感染性胃腸炎の患者がツイートした際、周りの人間のうち1人以上が感染性胃腸炎に患っている可能性が高い。そこで、近い距離に患者がいるとき陽性と判断する。距離がわからない場合は陰性とみなす。リツイートタイプのツイートは患者によるものではないとみなし、陰性を付与する。

条件2: 時制は現在か否か

症状の時制がツイートした時点に近いツイートを陽性とみなす。ここで、症状の時制は1日前つまり“昨日”を指すツイートまでを陽性とする。

表1. クロールした陽性ツイート例

キーワード	本文
悪心	“嘔吐でダウン”
下痢	“体調は良くなったけど下痢が止まりません感染性胃腸炎つらし”
寒気	“寒気するし気持ち悪い..(´・ω・`)”

構築したコーパスを用いて、SVMベースの分類器の訓練を行った。コーパスは、文とラベル(陽性もしくは陰性)の組から構成されている。コーパスに関するより正確な情報については、以前の報告書に示す。

最後に、地域ごとの流行度合いの把握のために、ツイートを都道府県ごとに分類をする。ツイートに付与された2種類のメタデータに基づいて地域を決定する。

1. GPS情報: ツイートには、Twitter ユーザがロケーション機能を許可している場合、GPS情報が付与されている。しかし、ほとんどのユーザはプライバシー上の理由でこの機能をOFFにしている。収集した

ツイートのうち GPS 情報が含まれているものは全体の内 0.46% (=35, 635/7, 666, 201) であった。

- プロフィール情報: いくつかのユーザのプロフィールには住所が含まれているものがある。それらのユーザは含まれている住所に住んでいるとみなした。プロフィールによる位置情報が含まれているツイートは 26.2% (=2, 010, 605/7, 666, 201)。我々はこのとき、地名を明確にするため Google Map の Geocoding service を用いた。具体的には、Twitter ユーザの位置名を Google Maps に送信し、受信した結果を JSON 形式で取得した。我々は、これらの受信した結果を解析し、位置情報を取得するため Python による構文解析を行った。

3.2 多重線形回帰による患者数予測

患者数と陽性のツイート数を用いて地域ごとに罹患率を推定する。このとき、患者数には国立感染症研究所 (NIID) から報告された患者数を用いた。NIID では各都道府県での患者数が毎週報告されている。地域の人口の偏りを解消するため、本研究では地域内の罹患率 (IR) を以下のように 1 週間ごとに算出した。

$$IR(a, t) = \frac{pat_{a,t}}{pop_a} \times 10^k \quad (1)$$

このとき、 $pat_{a,t}$ は、任意の都道府県 a の t 週目における総患者数、 pop_a は都道府県 a の総務省より発表された総人口。 k は値を修正するための定数である。今回は、 k を 5 に設定する。

次に、多重線形回帰を適用し、罹患率と推定患者数 y_a との間の説明変数を算出する。

$$y_a = b_a^{s1} x_a^{s1} + b_a^{s2} x_a^{s2} + b_a^{s3} x_a^{s3} + b_p, \quad (2)$$

このとき、 y_a は都道府県 a の予測罹患率、 x_a^s は症状 s の陽性ツイート数、 b_a^s 、 b_p は説明変数とする。

4 評価

4.1 実験

3.2 節で得られた説明変数に基づき、陽性ツイート数から罹患率を推定した。

データセットはトレーニングセットとテストセットで構成される。表 2 にコーパスの内訳を示す。トレーニングのため、2016 年 3 月 19 日から 2016 年 12 月 31 日までの 52 週間、計 1, 720, 325 のツイートを使用した。

表 2. データセット内訳

キーワード	トレーニング	テスト
悪心	560, 620 (53%)	594, 443 (50%)
寒気	378, 652 (37%)	498, 748 (35%)
下痢	781, 053 (74%)	493, 693 (69%)
総数	1, 720, 325 (34%)	1, 586, 884 (51%)

式 1 で定義した罹患率 (IR) を算出した。このとき、2017 年 1 月 1 日から 2017 年 4 月 19 日までの 16 週間でテスト

セットとし、IR を算出した。

4.3 結果

推定値と報告数が高い相関を示した例である長野県での推定結果図 2(a) を示す。東京都での推定結果を示す図 2(b) では、推定値が報告値と大きく異なることを示しており、我々のアプローチの弱点を示していると言える。図 2(a) と図 2(b) の違いは、流行規模の影響であると考えられる。図 2(a) では、罹患率は 1 つの大きなピークが見られる (罹患率 20~30 程度)。これに対して、図 2(b) では実際の患者数の増加はわずかしかない (15 週で罹患率 13~15 程度)。これらの結果から、小規模な流行の推定は困難であることがわかった。5 章では推定性能を落とす原因の議論を行う。

5 考察

5.1 流行規模と推定性能

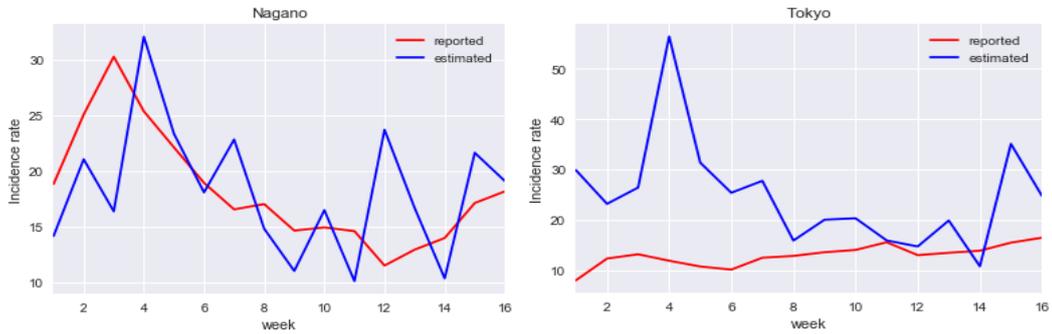
実験結果より、Twitter を用いた推定は小さな増減の悪影響を受けることが多く、推定性能は全体的に低いことがわかった。しかし、大規模な流行の場合、ソーシャルセンサは正常に動作することがわかった。それでは、どの程度の規模の流行であれば、ソーシャルセンサで捉えられるか確認をするため、推定性能 (Estimation performance 以下では EP と表す) と流行規模 (Event scale 以下では ES と表す。) の関係を調べた。以下にこれらの指標の定義を示す。

推定性能 (EP): ソーシャルセンサがどの程度正確に流行を推定できるか確認する。 $IR_{reported}$ と $IR_{estimated}$ との相関係数を推定性能 (EP) と定義する。

流行規模 (ES): 基本的に EP は流行規模 (ES) が大きいときに高い値を示す。 ES を以下のように定義する。

$$ES = \max(IR(a, t)) - \min(IR(a, t)), \quad (3)$$

このとき、 T は対象とするのセットの時系列を示す。 $IR(a, t)$ は、週目 t における IR を示す変数である。セットについては、4 週間ごとにテストセットを分割し (1 つのセット)、 IR を求めた。 EP と ES の相関を図 3 に示す。図 3 から EP と ES の間の相関は極めて低いことがわかる。この結果は、 EP と ES との間には全体的な相関が存在しないことを示している。



(a) 例；長野県

(b) 例；東京都

図 2. 結果の例として(a)長野県と(b)東京都を示す. このとき, X軸は時間(週単位)を示し, Y軸は罹患率(IR)を示している. 赤い線は実際に報告された罹患率, 青い線は推定された罹患率を示す.

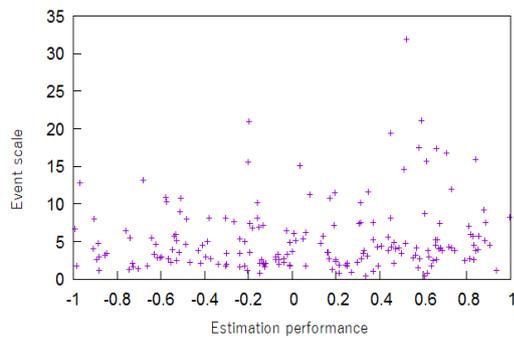


図 3. 全体での ES と EP の相関

5.2 流行推移に基づく考察

推定性能は流行規模(ES)だけでなく, 流行推移(患者数推移)にも影響を受ける. 流行の段階で分れる流行の週類(開始時と終了時)を考慮した.

1. 上昇 \nearrow は, 目的のセット内でIRが増加する流行の段階
2. 下降 \searrow は, 目的のセット内でIRが減少する流行の段階
3. 山 \wedge は, IRの最大値が観測される流行のピーク
4. 谷 \vee は, 2つの流行ピークの間

これらの定義の詳細を表3に示す. この表は, $IR_b - IR_i > 0$ と $IR_i - IR_e > 0$ の推移によって流行の推移を示す.

結果は, 表4に示しており, 各流行推移の流行規模と推定精度の相関を示す. 表に示すように, 性能は各推移で異なる. 下降期には高い相関を示した($r = 0.305$). 対照的に, 谷期では低い相関を示した(相関0未満).

表 3: 流行推移のタイプ一覧: それぞれのタイプは IR_b , IR_i と IR_e の3つのパラメータで区別される. IR_b は対象セットの先頭であるIRを表す. IR_i は, 最初と終わりを除くすべての週の平均を表す. セットサイズは4週間で設定する. したがって, IR_i は2週間目と3週間目の平均のIRである. IR_e はターゲットウィンドウの最後のIRである.

流行推移	$IR_b - IR_i$	$IR_i - IR_e$	サンプル数
上昇 \nearrow	+	+	55
下降 \searrow	-	-	45
山 \wedge	+	-	56
谷 \vee	-	+	32

表 4. 各推移の EP と ES の相関

流行推移	相関
上昇 \nearrow	0.098
下降 \searrow	0.305
山 \wedge	-0.024
谷 \vee	0.058

5.3 地域人口に基づく考察

各地域内のツイート数は, 各地域の人口に比例する. したがって, EPは各地域の人口の影響を受けると考えられる. 我々は, 各ウィンドウを人口に基づいて4つのタイプに分類した. 以下に4つのタイプの詳細を示す.

超高人口地域 Super High population area (SHP)は250万人以上の人口を持つ地域

高人口地域 High population area (HP)は人口150万~250万人の地域

低人口地域 Low population area (LP)は人口が100万~150万人の地域

超低人口地域 Super Low population area (SLP)は人口が100万人以下の地域

表5にEPとESの各タイプの相関を示す. 表5より, 高

人口地域(150 万~200 万人) では、弱い相関が見られた ($r = 0.214$)。結果から、EPとESの相関は、各地域の人口が関係していることがわかる。

5.4 各タイプの組み合わせによる考察

上述のように、(1) 流行規模 (ES)、(2) 流行推移、(3) 人口という推定性能に影響を与える3つの要因を導入した。上記の知見を組み合わせ、(2) 流行推移4種類と(3)人口別4種類の組16種類ごとに(1)事象規模との相関係数を調べた。得られた16パターンを表6に示す。表6から、低人口地域で流行が大幅に減少したとき、推定性能が高くなる傾向があることがわかる。特に、この傾向は、低人口および超低人口地域における顕著にみられた。対照的に、谷期のとき相関は低いことがわかる。実用の観点からは、患者の増加または減少をとらえることが予防に寄与するため、流行推移が増加傾向にある期間における推定性能が重要である。図4は、高人口地域で患者数が増加傾向にあるときの推定性能と流行規模の関係を示しており、中程度の相関 ($r = 0.378$) を示している。この図から、流行規模が (Y 軸) で7より大きい場合、中程度の推定性能 (X軸で $r > 0.8$) が得られることがわかる。このとき、流行規模 = 7は、一カ月に150人以上の患者増加に相当する。この結果から、(低人口地域で)患者数の減少、(高人口区域で)150人以上の患者数が増加のとき信頼性の高い推定が可能といえる。

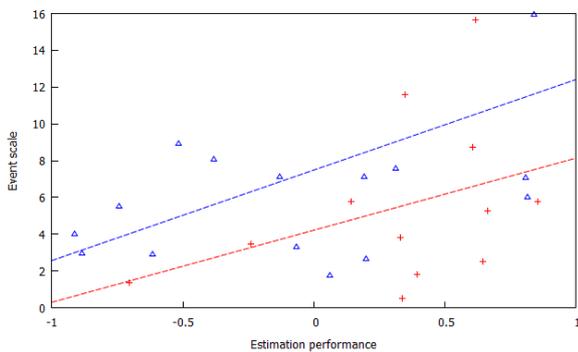


図4. タイプ別によるEPとESの相関; 赤い線(赤い十字でプロット)は上昇傾向における高人口地域、青い線(青い三角でプロット)は下降傾向における低人口地域を示す。

表6. タイプの組み合わせによるEPとESの相関

	超高人口 地域	高人口 地域	低人口 地域	超低人口 地域
上昇	0.255	0.378	0.01	-0.112
下降	-0.678	0.164	0.550**	0.538*
山	0.144	0.063	-0.394	0.411
谷	0.494	0.411	-0.409	0.179

*は相関が有意であるものを示す(**は $\rho < 0.05$, *は $\rho < 0.10$ を示す)。

5.5 実用化に向けて

今日まで、ソーシャルセンサは、様々な流行検出の可能性を実証されてきた。しかしながら、実用化にはほとんど至っていない。一つの理由として、ソーシャルセンサの信頼性の欠如が挙げられる。言い換えれば、ソーシャルセンサに基づく情報を完全に信頼することができない。本研究の結果は、ソーシャルセンサの推定性能は流行規模に関係していることを示した。この特性は次の二つのユースケースに対して重要な情報であると考えられる。

1. 実際に大きな流行が発生した場合、システムは流行の手がかりの検出を行う。
2. 対照的に、システムによる推定が流行を検出していない場合、少なくとも現状は流行していないと考えられる。

実用の観点から、これらの特徴が重要である。これらの結果を踏まえ、流行監視サービス(図5)を開発している。近い将来、システムの動作結果の事例を報告したいと考える。



図5. サーベイランスシステム

6 おわりに

本稿では、小規模および非季節性の疾患事例の検出における試みを記載した。この方法は、症状に関連する単語の出現頻度を用いる。2017年1月から2017年4月までの日本47都道府県を対象とした実験結果から、小規模な事象の検出は困難であることを示唆している (ESとEPの相関=0.13)。総合的な精度の結果は低いが、流行規模(患者数の変化)と検出性能の高さには相関がある高人口地域における流行段階では ($r = 0.38$ の相関比を示す)。今回の発見は以下の2つのユースケースに対して重要な情報になると考えられる。

1. 実際に大きな流行が発生した場合、システムは流行の手がかりの検出を行う。
2. 対照的に、システムによる推定が流行を検出していない場合、少なくとも現状は流行していないと考えられる。

これらの特性は、公共の安全を保護する上で重要である。

参考文献

- 1) Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. 2011. Twitter catches the flu: Detecting influenza epidemics using twitter. In Proceedings of the International Conference on Empirical Methods in Natural Language Processing (EMNLP). pages 1568- 1576.
- 2) Tammy L Stuart Chester, Marsha Taylor, Jat Sandhu, Sara Forsting, Andrea Ellis, Rob Stirling, and Eleni Galanis. 2011. Use of a web forum and an online questionnaire in the detection and investigation of an outbreak. *Online Journal of Public Health Informatics* 3(1).
- 3) Corinna Cortes and Vladimir Vapnik. 1995. Support vector networks. *Machine Learning* 20(3):273-297.
- 4) Aron Culotta. 2013. Lightweight methods to estimate influenza rates and alcohol sales volume from twitter messages. *Lang. Resour. Eval.* 47(1):217-238. <http://dx.doi.org/10.1007/s10579-012-9185-0>.
- 5) Ernesto Diaz-Aviles and Avar'e Stewart. 2012. Tracking twitter for epidemic intelligence: Case study: Ehec/hus outbreak in germany, 2011. In Proceedings of the Fourth Annual ACM Web Science Conference. *WebSci '12*, pages 82-85.
- 6) J. Espino, W. Hogan, and M. Wagner. 2003. Telephone triage: A timely data source for surveillance of influenza-like diseases. In Proceedings of the AMIA Annual Symposium. pages 215-219.
- 7) Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. 2009. Detecting influenza epidemics using search engine query data. *Nature* 457(7232):1012- 1014.
- 8) Hayate Iso, Shoko Wakamiya, and Eiji Aramaki. 2016. Forecasting word model: Twitter-based influenza surveillance and prediction. In Proceedings of International Conference on Computational Linguistics (COLING). Osaka, Japan, pages 76-86. <https://www.aclweb.org/anthology/C/C16/C161008.pdf>.
- 9) Thorsten Joachims. 1998. Text categorization with support vector machines: Learning with many relevant features. *Machine Learning: ECML-98* pages 137-142.
- 10) Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. 2004. Applying conditional random fields to japanese morphological analysis. In EMNLP. volume 4, pages 230-237.
- 11) Vasileios Lamos and Nello Cristianini. 2010. Tracking the flu pandemic by monitoring the social web. In Proceedings of International Workshop on Cognitive Information Processing (CIP). pages 411-416.
- 12) Vasileios Lamos, Andrew C Miller, Steve Crossan, and Christian Stefansen. 2015. Advances in nowcasting influenza-like illness rates using search query logs. *Scientific Reports* 5.
- 13) Vasileios Lamos, Bin Zou, and Ingemar Johansson Cox. 2017. Enhancing feature selection using word embeddings: The case of flu surveillance. In Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, WWW '17, pages 695-704. <https://doi.org/10.1145/3038912.3052622>.
- 14) S. Magruder. 2003. Evaluation of over-the-counter pharmaceutical sales as a possible early warning indicator of human disease. In Johns Hopkins University APL Technical Digest (24).
- 15) Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. pages 3111-3119.
- 16) Michael J Paul, Mark Dredze, and David Broniatowski. 2014. Twitter improves influenza forecasting. *PLoS Currents Outbreaks* .
- 17) Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes twitter users: Real-time event detection by social sensors. In Proceedings of the International Conference on World Wide Web (WWW). pages 851-860.
- 18) Shoko Wakamiya, Yukiko Kawai, and Eiji Aramaki. 2016. After the boom no one tweets: microblog based influenza detection incorporating indirect information. In Proceedings of the Sixth International Conference on Emerging Databases: Technologies, Applications, and Theory. ACM, pages 17-25.
- 19) Qian Zhang, Nicola Perra, Daniela Perrotta, Michele Tizzoni, Daniela Paolotti, and Alessandro Vespignani. 2017. Forecasting seasonal influenza fusing digital indicators and a mechanistic disease model. In Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, WWW '17, pages 311-319. <https://doi.org/10.1145/3038912.3052678>.