

Twitter Catches the Flu: 事実性判定を用いた インフルエンザ流行予測

荒牧英治[†] 増川佐知子[†] 森田瑞樹^{††}

近年の Twitter の普及とともに、そこから有益な情報を抽出する研究の需要が高まっている。本研究ではインフルエンザ流行予測の問題に焦点をあてる。まず、Twitter からインフルエンザに関連した発言を抽出する。次に、SVM を用いた分類器で、その発言者が実際にインフルエンザにかかっているかどうかを判定する。実験では、厚生省報告の症例数と比較し相関係数 0.89 の精度で流行予測することができ、提案手法の有効性を示した。特に、流行の開始時期においては、相関係数は 0.97 と高く、最先端の手法である Google Flu トレンドと同等の精度が得られた。本研究によって、Twitter 上のテキストが現実の疾病状態を反映していること、また、言語処理技術によって、その情報が抽出可能であることを示した。

Twitter Catches the Flu: Detecting Influenza Epidemics using Factuality Analysis

Eiji ARAMAKI[†], Sachiko MASKAWA[†], and Mizuki MORITA^{††}

With the recent rise in popularity and size of social media, there is a growing need for systems that can extract useful information from this amount of data. We address the problem of detecting influenza epidemics. First, the proposed system extracts the influenza related tweets using Twitter API. Then, the SVM based classifier extracts only tweets that include real influenza patients. The experiments demonstrated the feasibility of the proposed approach (0.89 of the correlation to the gold standard). Especially in the rise of the influenza (early stage of the epidemics), the proposed method shows the high accuracy (0.97 correlation), which is the same accuracy to the state-of-the-art method. This paper demonstrated that Twitter texts reflect the real world, and Natural Language Processing (NLP) technique could extract the information.

1. はじめに

近年、Facebook や Twitter などのマイクロブログが注目を集めている。本研究では Twitter に焦点をあてる。Twitter は、個々のユーザが「ツイート」(以下、単に**発言**とよぶ)と呼称される短文を投稿し閲覧できるコミュニケーション・サービスである。現在、1.2 億ユーザが参加しており、550 万ツイートが毎日やりとりされている(2011 年 3 月現在)。Twitter は、潜在的に多くのアプリケーションのための情報リソースとなる可能性を秘めており、Huberman 等[1]による人間関係の解析、Boyd 等[2]によるコミュニケーション活動の解析、Sakaki 等[3]による地震の探知など様々な活用法が提案されてきた。本研究では、インフルエンザの流行予測を題材とする。Twitter を利用したインフルエンザの流行予測には次の 2 つの利点がある

- **【大規模】** インフルエンザを含んだ発言は平均 1000 発言/日を超えている(Nov. 2008- Oct. 2009)。このデータのボリュームは、これまでの調査手法、例えば、本邦における医療機関の定点観測の集計を圧倒する大規模な情報収集を可能とする。
- **【即時性】** ユーザの情報を直接収集するため、これまでにない早い速度での情報収集が可能である。早期発見が重視されるインフルエンザ流行予測においては即時性が極めて重要な性質である。

以上のように、Twitter を用いたインフルエンザ流行予測は利点も多いもののその実現にあたっては、以下のような発言から不正確な情報を収集してしまう恐れがある：

- 頭痛... **インフルエンザ**かもしれない
- 厚生労働省は H1N1 **亜型インフルエンザ**ウイルスが現在、全国に広まりつつあると発表した。
- もしかして**インフルエンザ**じゃない？

上記の発言はすべて「インフルエンザ」という語を含んでいるものの、「インフルエンザにかかった人物が存在すること」を意味していない。このような発言を本稿ではイ

[†]東京大学知の構造化センター
Center for Knowledge Structuring, University of Tokyo.

^{††}独立行政法人医薬基盤研究所
The National Institute of Biomedical Innovation

a <http://www.facebook.com/>

b <http://twitter.com/>

インフルエンザ陰性発言と呼ぶことにする（逆に、インフルエンザにかかった人物が存在する場合をインフルエンザ陽性発言と呼ぶ）。本研究の実験データでは単語「インフルエンザ」を含む発言のうち 42% がインフルエンザ陰性発言であった。このような陰性発言が多いと、発言数と流行予測結果の間にずれがうまれ、不正確な値が得られてしまう。

そこで、本研究ではインフルエンザ陰性発言を取り除くために、機械学習による分類器を用いる。まず、発言に陽性／陰性ラベルを付加したコーパスを構築する。次に、そのコーパスを用いて Support Vector Machine (SVM) (Cortes and Vapnik[4]) による文分類器を構築する。実験の結果、陽性陰性の分類において F 値 0.76 が得られ、一定の精度でインフルエンザ陽性発言だけを抽出ことが可能であった。また、インフルエンザ陽性発言の相対頻度を単位時間ごとに集計すると、現実のインフルエンザ流行（定点当たりの患者数）との相関係数が 0.89 という高い値を示し、流行予測が可能であることを示した。

特に、流行の開始時期においては、相関係数 0.97 と高い精度が得られ、最先端の手法である Google Flu トレンドの精度 (Ginsberg et al.[5]) を上回った。

本論文のポイントは次の 2 点である：

- (1) SVM による分類器で Twitter 上の発言からインフルエンザ陰性発言を除去できることを示す。
- (2) インフルエンザ陰性発言を除去することにより、高い精度でインフルエンザ流行を予測可能であることを示す。

2. インフルエンザ流行予測

インフルエンザの流行予測は、事前の十分なワクチン準備が必須の対応となるため、世界各国で共通の課題として対策が行われている。これまでは国からの主導による集計が中心であったが (2.1 節)、現在では、ITC 技術を用いた大規模な調査方法が検討されている (2.2 節)。

2.1 これまでの調査方法

多くの国は、自国の感染症調査のための機関を備えている。例えば、アメリカでは Centers for Disease Control and Prevention (CDC), E.U は EU influenza Surveillance Scheme (EISS) を運営している。本邦では、国立感染症研究所 (Infection Disease Surveillance Center; IDSC) がインフルエンザの流行予測を行っている。これらの機関は主としてウイルスの遺伝子解析、国民の抗体保有状況や協力体制にある医療機関からの報告に頼って集計を行っている。例えば、IDSC は全国約 5,000 の診療所から情報を集め、集

計結果を発表している。ただし、この集計には時間がかかり、1 週間前の流行状況が発表されるため、発表時にはすでに流行入りしている可能性があり、より迅速な情報収集への期待が高まっている。

2.2 新しい調査方法

より早くインフルエンザの流行を捉えるため様々な手法が提案されている。Espino 等 [6] は電話トリアージ・アドバイス（電話を通じて医療アドバイスをこなう公共サービス）に注目し、電話コールとインフルエンザの関係に相関があり、インフルエンザ流行予測において有益な情報源になることを明らかにした。

Magruder [7] は、インフルエンザ市販薬の販売量 (over-the-counter drug sales) に注目した。ただし、インフルエンザ市販薬は一部の国でしか販売されておらず（例えば、本邦ではインフルエンザ薬の購入には処方が必要となる）、常に有効な手段ではない。

近年、もっとも注目されている手法は Ginsberg 等 [5] の Google の web 検索クエリを用いた手法である。彼らはインフルエンザ流行と相関のある検索クエリ（相関係数の高い上位 50 語）を調査し、それらをモニタリングすることで、インフルエンザの予測を行っている。彼らの予測は、アメリカの CDC 報告との相関係数 0.97 (min=0.92; max=0.99) という高い値を報告している。同様のアイデアにもとづいた研究は他にも報告されている。Polgreen 等は同様の手法で Yahoo! のクエリを用いた (Polgreen et al. [8])。Hulth 等はスイスの web 検索会社のクエリを用いた (Hulth et al. [9])、Johnson 等は健康サイト「HealthLink」の web ページ閲覧のアクセスログを用いた (Johnson et al. [10])。

上記に紹介した手法はそれぞれ異なった情報を用いているものの、患者の行動を直接調査するという観点からは同様のアプローチであるとみなせる。当然、患者の行動が分かれば多くの情報を即時的に収集することが可能となるが、これらを行える機関は限られているという問題がある。例えば、web 検索クエリを用いることができる機関は Google, Yahoo! や Microsoft といったいくつかのサービス・プロバイダに限定される。そこで、本研究では Twitter という web 検索クエリに比べ利用が容易な情報源を用いて、疾患状態の把握を目指す。

3. インフルエンザ陽性-陰性コーパス

1 章にて述べたように、Twitter を用いてインフルエンザ流行予測を行なうためには、インフルエンザ陰性発言を取り除く必要がある。これを行なうために、本研究ではインフルエンザ・コーパスを構築し (3 章)、これを用いて SVM によるインフルエンザ陽性陰性発言の分類器をつくる (4 章)。ここでいうコーパスは発言そのものとインフルエンザ陽性／陰性のラベルのペアからなり、次の手順で構築した。

c <http://sankei.jp.msn.com/life/news/110112/bdy11011222430044-n1.htm>

3.1 データ収集

まず、2008年11月から2010年7月にかけてTwitter APIを用いて30億発言を収集した。収集結果を図1に示す。次に、「インフルエンザ」という語を含む発言を収集した(40万発言)。さらにこのデータを2つの領域に分割した。

- **【トレーニング・データ】** 2008年11月から無作為に5000発言。このデータに人手でアノテーションを行いインフルエンザ陽性-陰性コーパスとした。
- **【テスト・データ】** 残りのデータ。インフルエンザ推定での実験に用いた。

3.2 陽性-陰性アノテーション

それぞれの発言についてアノテーターが陽性または陰性のラベルを付与する。このラベル付けにおいて、陽性は次の2つの基準を満たすものとした。

- **【条件1:命題的的確性】** 発言者または発言者の付近の人物のインフルエンザについて話題になっている必要がある。ここでいう「発言者の付近」とは発言者と同じ都道府県であることが伺える場合とした。もし、発言者とインフルエンザにかかった人物との距離が不明な場合は、陰性とみなした。同様に、RT(リツイート；他人の発言をそのまま引用する発言)は陰性とした。
- **【条件2:時制/モダリティ的的確性】** 時制については、現在または近い過去であることを条件とした。ここでいう「近い過去」とは24時間以内とした。例えば、「昨日」などもこの条件を満たす。モダリティについては、仮定法や疑問文などインフルエンザの事実を損なう種類のモダリティを除いた。ここでいう事実性を損なうモダリティとは法(仮定法)、表現類型(疑問文、命令文)、価値判断(必要、許可)とした。詳細な基準についてはアノテーション・ガイドラインを参照のことd。

d アノテーション・ガイドライン <http://mednlp.jp/~aramaki/KAZEMIRU/>

陽性(+1)／ 陰性(-1)	発言 (ツイート)	判断理由
+1	インフルエンザが研究室で流行っている	
+1	インフルエンザだん	
+1	インフルエンザ発症(>_<)最近は薬も1回だけでいい なんてビックリ!	
+1	まさかのインフルエンザ発症…早く治りますように	
+1	出社するとインフルエンザをうつされにいくような もんだ	
+1	インフル。昨日よりひどい熱、	
-1	去年はひどいインフルエンザで参加できなかった	過去時制
-1	インフルエンザになると耳鼻科では意味がない	一般知識
-1	インフルにもかかってしまえ	命令文
-1	インフルエンザでもいい	価値判断
-1	インフルエンザだとしても、熱がないのが不思議	仮定法

表1: インフルエンザ・コーパス。

「+1」は陽性を示す。「-1」は陰性を示す。

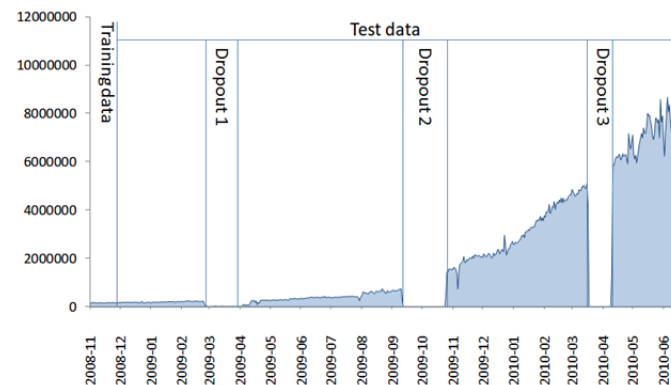


図1: Twitter クロール・データ

Twitter API の仕様変更のため3箇所のドロップアウト (Dropout1..3) が存在する。ドロップアウト期間は、5章にて後述する実験より除いた。

4 インフルエンザ陽性-陰性分類器

3章にて述べたコーパスを用いて、入力された発言がインフルエンザ陽性か陰性を判定する分類器を構築する。このタスクは、スパムメール・フィルタリングや評価表現分析といった文を分類するタスクと類似している。本研究では、これらの研究で一般的な手法である注目している語の周辺文脈を素性とした機械学習による手法を用いた(図2)。周辺文脈をどれだけの大きさみるか(ウィンドウ・サイズ)(4.1節)やどんな機械学習手法がよいか(4.2節)は予備実験によって調査した。予備実験は前章のトレーニング・データを10分割交差検定することによって行った。

4.1 素性のウィンドウ・サイズ

パフォーマンスはウィンドウ・サイズに依存する。図3にウィンドウ・サイズと精度(F値)の関係を示す。最もよいパフォーマンスは左右両方のコンテキストを6語までみる設定(BOTH=6設定)で得られたため、以降の実験ではこの値を用いる。インフルエンザを判定するためには、発言すべての語を用いる(BOTH=∞)を用いるのではなく、周辺に絞った方がよいと分かる。

4.2 機械学習手法

精度(F値)と速度(学習時間)という2つの観点から、様々な機械学習手法を比較した(表2)。結果が多項カーネルを用いたSVMがもっとも高い精度と二番目に早い速度を示し、これらを考慮し、以降の実験ではSVMを用いることとした。

5 実験

国立感染症研究所IDSCのデータを用いた実験を行った。

5.1 比較手法

提案手法と精度を比較するため以下の手法を比較した。

- **TWEET-SVM**: 提案手法。
- **TWEET-RAW**: 単なる語の出現頻度によるベースライン。「インフルエンザ」を含んだ発言の相対頻度を出力する。
- **DRUG**: 風邪薬(総合感冒薬)の販売量。厚生省の調査による。
- **GOOGLE**: Google flu trendの日本語バージョンe。Google web検索のクエリを用いている(Ginsberg et al.[5])。

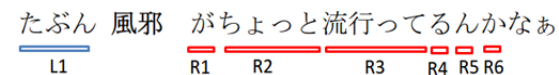


図2: 注目している語(「風邪」)の周辺文脈を素性とする。

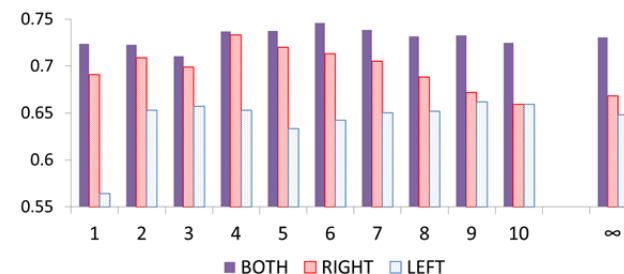


図3: ウィンドウ・サイズと精度(F値)。

RIGHTは右コンテキストのみを用いた場合、LEFTは左コンテキストのみを用いた場合、BOTHは両方のコンテキストを用いた場合を示す。横軸の数字はウィンドウ・サイズを示す。∞はその方向のすべてのコンテキストを用いることを示す。例えば、BOTH∞は注目している語以外のすべての語を用いる。

機械学習手法	F値	学習時間(秒)
AdaBoost (Freund [11])	0.592	40.192
Bagging (Breiman [12])	0.739	530.310
Decision Tree (Quinlan[13])	0.698	239.446
Logistic Regression	0.729	696.704
Naive Bayes	0.741	7.383
Nearest Neighbor	0.695	22.441
Random Forest (Breiman [14])	0.729	38.683
SVM (RBF kernel) (Cortes and Vapnik [4])	0.738	92.723
SVM (polynomial kernel; d=2)	0.756	13.256

表2: 機械学習手法とパフォーマンス (F値と学習時間)

e <http://www.google.org/flutrends>

5.2 正解データとテストセット

正解データ (gold standard) として IDSC から毎週報告されているインフルエンザ定点当たりの患者数を用いた。これは、厚生労働省・感染症サーベイランス事業により、全国約 5,000 のインフルエンザ定点医療機関を受診した患者数が週ごとに報告されたものである。これらを以下の 4 つの期間に分けて、各手法で予測を行った。

- Season I (2008 冬)
- Season II: (2009 夏) (過熱報道時 (パンデミック騒動時))
- Season III: (2009 冬)
- Season IV: (2010 夏)

より詳細な評価を行なうために、インフルエンザが流行する冬は、流行前 (before peak) と流行後 (after peak) の 2 つの期間に細分割した。分割にあたってはシーズンでもっとも報告が多かった時点を境界にそれ以前を流行前、それ以降を流行後とみなした。

【過熱報道時】 実験データの期間中、Season II と Season III の before peak は新型インフルエンザ (swine flu; H1N1 flu) に関する報道が過熱していた時期である (図 4)。この時期は例外的な現象であるので、この期間を**過熱報道時**と呼び、これを個別に精度評価した。



図 4: 2009 年 6 月の CNN ニュース (左図) と 2009 年 4 月の読売オンラインニュースアーカイブ (右図)。

79 週					
Season I 2008/11/9- 2009/4/5		Season II 2009/4/12- 2009/7/5	Season III 2009/7/12- 2010/2/14		Season IV 2010/2/21- 2010/7/4
22 週		13 週	26 週		18 週
流行前 2008/11/9- 2009/1/25	流行後 2009/2/1- 2009/4/5		流行前 2009/7/12- 2009/11/29	流行後 2009/12/6- 2010/2/14	
12 週	10 週		15 週	11 週	
平常時		過熱報道時		平常時	

表 3: テストセットのトラックと期間 (週)

		TWEET-RAW	TWEET-SVM	DRUG	GOOGLE
		(提案手法)			
過熱報道時		0.001	0.060	0.844	<u>0.918</u>
平常時		0.831	<u>0.890</u>	0.308	0.847
		0.683	0.816	-0.208	<u>0.817</u>
Season I	流行前	0.914	<u>0.974</u>	-0.155	0.962
	流行後	0.952	0.955	0.557	<u>0.959</u>
Season II		-0.009	-0.018	<u>0.406</u>	0.232
		0.382	0.474	0.684	<u>0.881</u>
Season III	流行前	0.390	0.474	0.919	<u>0.924</u>
	流行後	<u>0.960</u>	0.944	0.364	0.936
Season IV		0.39	0.957	0.130	<u>0.976</u>

表 4: 実験結果 (相関係数)。

太線の数字は正解データと有意に相関している値を示す (p=0.05)。下線付きの数字は該当期間で最も高い相関係数を示す。

5.3 評価方法

評価方法は正解データと各手法の値の相関係数（ピアソンの相関係数）を用い、予測のよさを測った。

5.4 結果

結果を表 4 に示す。平常時においては、提案手法はもっとも高い精度を示した（相関係数 0.890）。これは、GOOGLE の精度と同等で、提案手法の妥当性を示すと考えられる。しかしながら、過熱報道時においては、低い精度を示し、ニュースによるバイアスを受けていることが分かる。

6 議論

6.1 SVM による分類は精度に貢献する

多くの期間で、提案手法(TWEET-SVM)は「インフルエンザ」の頻度を数える手法(TWEET-RAW)よりも高い精度を示した。精度の平均増加ポイントは 0.196 (max 0.56; min-0.009)であり、期間によっては有意に精度を高めている。この結果は言語処理技術を用いるアプローチの妥当性を示している。現時点では陽性/陰性の分類は F 値 76% にとどまるが、将来的により高精度な分類が可能となれば、言語処理を用いたアプローチの優位性はより高まるであろう。

6.2 過熱報道時での影響

ほぼすべての手法は過熱報道時において精度を下げている。特に Twitter ベースの手法は、劇的に精度が下がっている。これは Twitter が他の手法と比較して、報道のバイアスを受けやすいことを示している。この理由の一つとして考えられるのは、Twitter は検索クエリなどと違ってコミュニケーションツールであり、そこでの発言は他の参加者に影響を与えうるからであろう。この現象に対処することができれば今後、より高精度の予測が可能となるであろう。

6.3 実用的観点からみた Twitter のインフルエンザ早期発見の長所と弱点

実用的な観点からはいかに正確にインフルエンザの流行を事前に感知するかが重要となる。この精度は、流行前のデータ（Season I 流行前と Season III 流行前）を調べれば分かる。両時期における各手法の値を図 5 にプロットした。

Season I (図 5 左)では提案手法 (TWEET-SVM) が最も高い精度（相関係数 0.974）をみせている。この値は、TWEET-RAW（相関係数 0.914）よりも高く、GOOGLE の精度（相関係数 0.962）と同等である（相関係数の有意差の検定）。これは web 検索クエリのような利用が困難なデータを用いなくとも Twitter のみで同等な精度が得られることを示しており、マクロブログの可能性を示唆するものである。

一方、Season III (図 5 右)は、過熱報道時であり、提案手法を含めたすべての手法の精度が低い。この過熱報道時において、4 月のインフルエンザ報道開始と 6 月の WHO

のパンデミック宣言という 2 回の出来事が大きくニュースで取り上げられた。ちょうどこれらに対応して Twitter ベースの手法(TWEET-SVM and TWEET-RAW)ではグラフ中に 2 つのピークを見ることが出来る。このことから Twitter はニュースへ鋭敏に反応していることが分かる。今後の課題として、このような時期においても安定した予測を行なう技術が求められる。

6.4 流行前の人間は過敏に反応する

図 6 に Season I における 2 つの手法による予測値 (TWEET-SVM と GOOGLE) と正解データ (gold standard) との相関を示した散布図を示す (流行前は「+」、流行後は「-」で示されている)。推定された値は基本的に、正解データと比例関係にあるものの、「+」が「-」よりも高い位置にある。つまり、同程度のインフルエンザ流行レベルにおいて、流行前の方が流行後よりも大きな反応を起こしていることが分かる。これは、流行り初めに人間が大きく反応することを示している。今後、この性質を陽に取り込んだモデルを構築することができれば、より高精度なインフルエンザ流行予測が可能となると考えられる。

7 関連研究

提案手法の中心となる技術は発言を陽性/陰性に分類する手法である。このタスクは negation (identification) と呼ばれる否定表現の有無やその範囲を捉えるタスクと類似している。このタスクは情報抽出、特に医療分野でのテキスト処理、で熱心に研究されており、否定表現抽出のアルゴリズム (正規表現ルールの集合) として NegEx (Chapman et al.[15]), Negfinder (Mutalik et al.[16])や ConText (Chapman et al.[17])が提案されている。また、機械学習による否定表現の抽出も研究されている(Elkin et al.[18]) (Huang and H.J. Lowe[19])。

これらの先行研究と提案手法の違いは、先行研究の negation が統語的な否定表現を捉えているのに対して、本研究ではインフルエンザの有無という観点からの判定であり、ある種の意味的な区分である点である。両者の判定の異なりを表 5 に示す。一般的に、意味的処理は難しいが、限られたドメインで、しかも、陽性/陰性という二値判定という限定された処理ならば、意味的な処理も可能であることを本研究は示している。

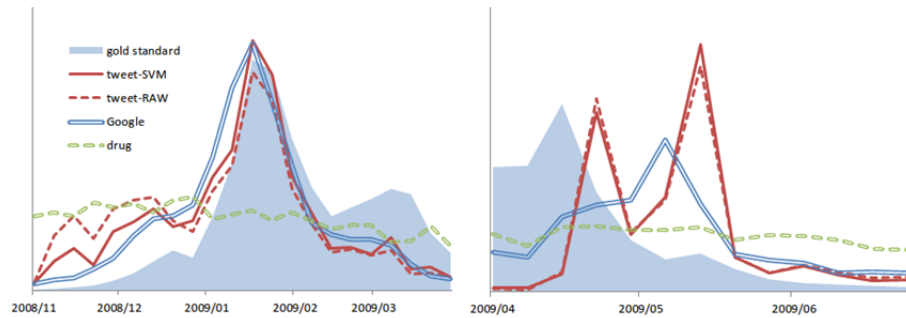


図 5: Season I (図左)と Season III (図右).

X 軸は日付を示している. Y 軸は各手法による予測と正解データを示す.

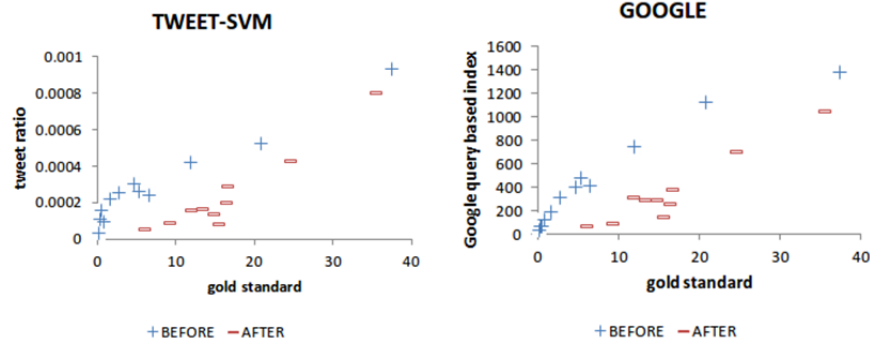


図 6: Season I における流行前と流行後の行動 (Tweet や web 検索)の感度.

正解データと各手法の予測の分布図 (図左 TWEET-SVM と図右 GOOGLE).

「+」は流行前を示す. 「-」は流行後を示す.

	Negation	インフルエンザ 陽性/陰性
インフルエンザなう	Positive	インフルエンザ陽性
インフルエンザじゃなかった	Negative	インフルエンザ陰性
インフルエンザについて勉強中	Positive	インフルエンザ陰性
インフルエンザが治らない...	Negative	インフルエンザ陽性

表 5: 先行研究の否定表現判定 (Negation) と本研究の陽性/陰性判定.

本研究のもう 1 つの特徴は, Twitter を用いた点である. 言語処理分野では Twitter は named entity recognition (NER) (Finin et al.[20]), sentiment analysis (Barbosa and Feng [21])など基礎的アプリケーションへの応用が多い. 本研究では, 我々の日々の生活に役立つアプリケーション構築に焦点を当てた.

8 おわりに

本研究では Twitter を用いたインフルエンザ流行予測を行った. SVM を用いたインフルエンザ陽性陰性発言の自動識別を行った結果 (F 値 0.76), 高い精度でインフルエンザ流行を予測できた (相関係数 0.89). 特に, 流行の開始時期においては, 相関係数 0.97 と高い精度を得ており, これは近年の ICT 技術を用いた予測の中で最も精度が高い報告である. ただし, インフルエンザ報道が過熱している時では安定した精度が出ず, 過熱報道時での取り扱いが今後の課題として残っている. 本研究は, Twitter テキストは実世界を反映した情報が流れていること, また, 言語処理によりその情報を一定の精度で抽出できることを示した.

謝辞 本研究は, JST 戦略的創造研究推進事業 (さきがけタイプ)「情報環境と人」及び, 科研費補助金(若手研究 A) (挑戦的萌芽)による. 本論文を書くにあたって有益な議論をいただいた東京大学医学部附属病院篠原恵美子氏, 知の構造化センター岡瑞起氏に謹んで感謝の意を表す.

参考文献

- 1) B. Huberman and D. R. F. Wu. Social networks that matter: Twitter under the microscope. First Monday, 14, 2009.
- 2) D. Boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In Proceedings of HICSS43, 2010.
- 3) T. Sakaki, M. Okazaki, Y. Matsuo, Earthquake shakes Twitter users: real-time event detection by

- social sen-sors, Proceedings of the 19th international conference on World Wide Web (WWW).
- 4) C. Cortes and V. Vapnik. Support vector networks. In *Machine Learning*, pp 273–297, 1995.
 - 5) J. Ginsberg, M.H. Mohebbi, R.S. Patel, and L. Brammer, Detecting influenza epidemics using search engine que-ry data, *Nature Vol. 457*, 19. 2009.
 - 6) J. Espino, W. Hogan and M. Wagner, Telephone triage: A timely data source for surveillance of influenza-like diseases. *AMIA: Annual Symposium Proceedings 215–219*, 2003.
 - 7) S. Magruder, Evaluation of over-the-counter pharma-ceutical sales as a possible early warning indicator of human disease. *Johns Hopkins University APL Tech-nical Digest 24*, 349–353, 2003.
 - 8) PM. Polgreen, Y. Chen, D.M. Pennock, F.D. Nelson, 2008. Using Internet Searches for Influenza Surveil-lance, *Clinical Infectious Diseases Volume 47 (11)* pp. 1443-1448.
 - 9) A. Hulth, G. Rydevik, A. Linde. 2009. Web Queries as a Source for Syndromic Surveillance. *PLoS ONE 4(2)*,
 - 10) HA. Johnson, MM. Wagner, WR. Hogan, W. Chapman, RT. Olszewski, J. Dowling, G. Barnas. 2004. Analysis of Web access logs for surveillance of influenza. *Stud Health Technol Inform. 107(Pt 2):1202-1206*.
 - 11) Y. Freund and R. Schapire, Experiments with a new boosting algorithm. In *Machine Learning Interna-tional Workshop*, pp148–156, Morgan Kaufmann Publishers, 1996.
 - 12) L. Breiman, Bagging predictors. *Machine learning*, 24(2):123–140, 1996.
 - 13) J. Quinlan. C4. 5: programs for machine learning. Mor-gan Kaufmann, 1993.
 - 14) L. Breiman, Random Forests. *Machine learning*, 45(1):5–32, 2001.
 - 15) W. Chapman, W. Bridewell, P. Hanbury, G.F. Cooper, and B. Buchanan, A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, 5:301-310, 2001.
 - 16) P.G. Mutalik, A. Deshpande, and P.M. Nadkarni, Use of general purpose negation detection to augment con-cept indexing of medical documents: A quantitative study using the umls. *Journal of the American Medi-cal Informatics Association*, 8(6):598-609, 2001.
 - 17) W. Chapman, J. Dowling and D. Chu, ConText: An algorithm for identifying contextual features from clinical text. *Biological, translational, and clinical language processing (BioNLP2007)*, pp.81–88, 2007.
 - 18) P.L. Elkin, S.H. Brown, B.A. Bauer, C.S. Husser, W. Carruth, Larry R. Bergstrom, and Dietlind L. Wahner Roedler. A controlled trial of automated classification of negation from clinical notes. *BMC Medical Informatics and Decision Making 5:13*, 2005.
 - 19) Y. Huang and H.J. Lowe, A novel hybrid approach to automated negation detection in clinical radiology re-ports. *Journal of the American Medical Informatics Association*, 14(3):304-311, 2007.
 - 20) T. Finin, W. Murnane, A. Karandikar, N. Keller, J. Mar-tineau, and M. Dredze. 2010. Annotating named entities in Twitter data with crowdsourcing. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk (CSLDAMT '10)*, 80-88, 2010.
 - 21) L. Barbosa and J. Feng. Robust Sentiment Detection on Twitter from Biased and Noisy Data. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, 2010.