

# ソーシャルセンサの地域性を考慮した インフルエンザ流行予測

若松 幸大<sup>†1,a)</sup> 清 雄一<sup>†1,b)</sup> 田原 康之<sup>†1,c)</sup> 大須賀 昭彦<sup>†1,d)</sup>

**概要:** 東京都ではインフルエンザ患者数の発表は東京都健康安全研究センターが行っており、定点医療機関からの報告を基に、インフルエンザ患者数を集計している。しかし実際にインフルエンザ患者が確認されてから一週間ほど遅れて情報が公開されている。つまり、インフルエンザの流行予測をすることで検疫、予防注射、抗ウイルスキャンペーン、病院リソースなどに役立つことができる。インフルエンザ流行予測の研究は様々行われているが、地域性を考慮したピンポイント予測というものは行われていない。そこで本論文では、リアルタイム性を考慮したピンポイントインフルエンザ予測をソーシャルセンサと LSTM を用いて行った。地域性が考慮されたインフルエンザ関連ツイートは決定係数が 0.9516 であり、ピンポイントインフルエンザ流行予測に役立つことが分かった。

## 1. はじめに

### 1.1 背景

インフルエンザは、重篤な病気を引き起こし、死に至るおそれのある急性呼吸器感染症である。日本では例年 1,000 万人近くがインフルエンザにかかり、インフルエンザ関連の合併症により約一万人が死亡する [1][2]。

東京都感染症情報センターではインフルエンザのデータの利用、公開までに 1~2 週間の遅延がある。リアルタイム性が欠如してどの程度流行しているかが 1~2 週間遅れてしか正確な情報がわからない。つまり、効果的なインフルエンザの予測と早期の発生検出は、検疫、予防注射、抗ウイルスキャンペーン病院リソースの管理に役立つ。

インフルエンザ予測の研究は多岐にわたるが、国レベルなど範囲の広いものを対象としたものが多い。しかし、インフルエンザは東京と大阪など離れた地域では流行時期が 1 週間から 2 週間ほど異なることがある (図 1)。図 1 は、横軸が週を表し、縦軸がインフルエンザ患者数を表す。また、図 2 のように東京の中でも区が違えば流行時期が 1 週~2 週ずれることが多々ある。つまり、インフルエンザ予測の対象範囲が広いものは特定の地域での正確な予測ができていとは言えない。特定の地域のインフルエンザ

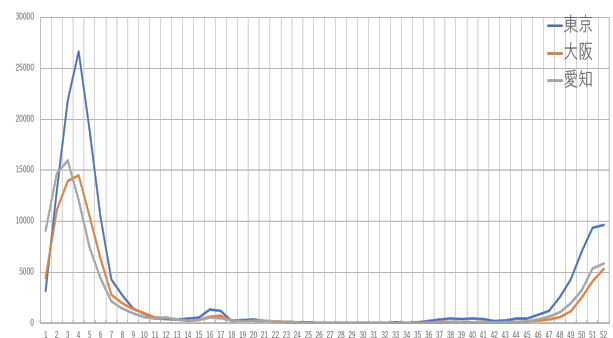


図 1: 2019 年の東京, 大阪, 愛知, のインフルエンザ患者推移

予測はより現実的な問題を解決できる。ワクチンの需要量を正確に測ることや、病院のリソース管理に特に役立つ。また、インフルエンザは飛沫感染、接触感染、空気感染によって感染するため、ピンポイントで予測することで個人に対して的確に予防を呼びかけることができる。

本研究ではピンポイントのインフルエンザの傾向を予測するために、Long short-term memory (LSTM) リカレントニューラルネットワークを採用する。インフルエンザは季節性の病気であるため気象情報に着目することでインフルエンザの予測に役立つことができる [6]。しかし、インフルエンザは寒いときに必ずしも発症するわけではない [7]。季節外れのインフルエンザが発生する可能性があり、気象情報だけではインフルエンザの流行検知に対処できない可能性がある。そこで、地域性に着目することでインフルエンザの傾向の予測精度を上げることを目指す。地域性を獲得するために Twitter を用いる。

<sup>†1</sup> 現在、電気通信大学 〒182-8585 東京都調布市調布ヶ丘 1-5-1  
Presently with The University of Electro-Communications  
Chohu, Tokyo 182-8585, Japan

a) wakamatsu.kodai@ohsuga.lab.uec.ac.jp

b) seiuny@uec.ac.jp

c) tahara@uec.ac.jp

d) ohsuga@uec.ac.jp

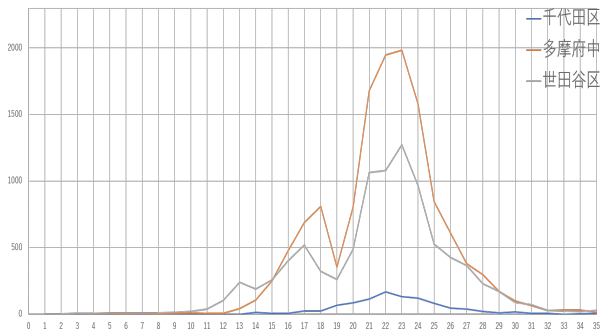


図 2: 2017 年 35 週から 2018 年 17 週のインフルエンザ患者推移

## 1.2 論文の構成

本論文の構成は次の通りである。2 章で関連研究について述べる。3 章では提案手法に関して、4 章では実験について述べる。5 章では考察、最後に 6 章で本論文の結論をまとめ、今後の展望を記す。

## 2. 関連研究

### 2.1 インフルエンザ流行予測

インフルエンザ予測の研究は様々な方法で行われている。大きく分けて 3 種類がある。一つ目は Susceptible-Infected-Recovered(SIR)[8], Susceptible-Infected-Recoverd-Susceptible(SIRS)[9], Susceptible-Exposed-Infected-Recovered(SEIR)[10] のようなコンパートメントモデル。これらのモデルは、感染症流行の数理モデルであり、Susceptible: 感染症に対して免疫を持たない者, Infected: 発症者, Recovered: 感染症から回復し免疫を獲得した者, Exposed: 感染症が潜伏期間中の者というふうに分け感染症がどのように伝播し、感染した人がどの程度の期間で発症し、重症化するのかといったプロセスを数式で記述した。

二つ目は自己回帰統合移動平均 (ARIMA)[11] や一般化自己回帰移動平均 (GARMA) [12] のような統計的及び時系列ベースの手法。ARIMA モデルには独立変数と従属変数の間に線形関係を想定している。実世界の関係は非線形であるため、モデルに組み込まれた仮定より複雑であるという欠点がある。また、GRAMA モデルにはパラメータの最適化に問題がある。

三つ目は機械学習を利用したインフルエンザ流行予測がある。近年では特に機械学習を用いたインフルエンザ予測の研究が行われている [3][4][6]。[3] ではインフルエンザ流行予測にツイートをを用い、Twitter のテキストが現実の疾病状態を反映していることが示された。また、[4] では Twitter の位置情報を利用することでインフルエンザ流行予測の精度を上げている。しかし、Twitter の位置情報付きツイートは全体のツイートの約 0.18% 程度である [5]。位置情報付きツイートの少なさから特定できない地域なども

多数あり、ピンポイント予測には向かない。[6] では LSTM の手法を導入し、外部変数として温度、気温、降水量、日照時間などの気象変数と周辺地域のインフルエンザ患者数を考慮することでインフルエンザの予測を行っている。

しかし、この手法ではピークの位置や観測値の急激な増加や減少に対応することが難しい。また、この研究では対象範囲をアメリカの複数の州を合わせたものを一つの地域としているため観測値の幅がそれほど大きくない。つまり、一つの州や州の中のある一つの地域のインフルエンザ患者数と研究で用いられているモデルが完全に一致しない可能性が考えられる。東京都の中で比べてみても図のようにインフルエンザの流行初期の時期やピークの位置がずれている。そこで今回は、調布・府中地域という小さな地域に限定し最適なモデルを提案することを目標とする。ツイートをを用いることで地域性を考慮し、インフルエンザ流行の初期の段階の予測、ピーク位置の予測精度の向上を目指す。

### 2.2 LSTM

LSTM は RNN の一種であり、長期の時間依存性も短期の時間依存性も学習できる手法である。

#### 2.2.1 LSTM セル

LSTM の手法自体は新しいものではなく、1997 年に初めて発表された [15]。その後、手法が改良され忘却ゲートが導入された [16]。RNN では、時間を深くさかのぼるほど勾配が消失してしまうという問題があった。そこで、LSTM ではネットワーク (ニューロン) の構造を変えるというアプローチによって勾配消失問題を対処した。また、忘却ゲートを取り入れることで、CEC (constant error carousel) の値を直接書き換えることができるようになった。LSTM セル図 3 は次のような式で成り立つ。入力を  $x(t)$ 、出力を  $h(t)$  とすると、

$$i(t) = \sigma(W_i x(t) + U_i h(t-1) + b_i) \quad (1)$$

$$f(t) = \sigma(W_f x(t) + U_f h(t-1) + b_f) \quad (2)$$

$$o(t) = \sigma(W_o x(t) + U_o h(t-1) + b_o) \quad (3)$$

$$c(t) = f(t) \odot c(t-1) + i(t) \odot a(t) \quad (4)$$

$$a(t) = f(W_a x(t) + U_a h(t-1) + b_a) \quad (5)$$

$$h(t) = o(t) \odot g(c(t)) \quad (6)$$

$i(t)$ ,  $f(t)$ ,  $o(t)$  はそれぞれ入力ゲート、忘却ゲート、出力ゲートを示している。 $a(t)$  は  $f$  によって活性化された値であり、 $c(t)$  は CEC の値である。 $W$  と  $U$  は重み行列を、 $b$  はバイアスペクトルを表している。また、 $\sigma$  は活性化関数を示す。

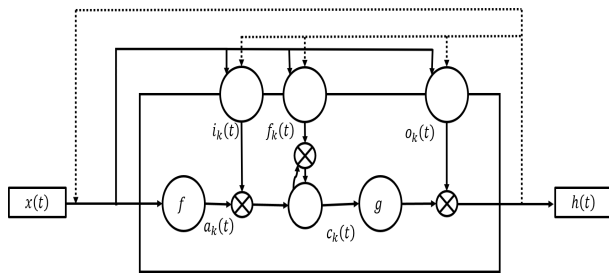


図 3: LSTM セル

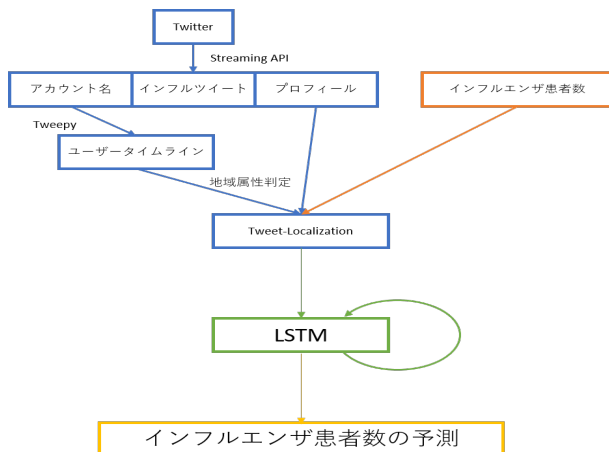


図 4: 提案手法の概略図

### 3. 提案手法

#### 3.1 提案手法の概要

既存研究では、インフルエンザ予測をする対象として国レベルなど対象範囲が広いものだけであった。そこで、本研究では調布・府中地域という粒度の小さい範囲に対してインフルエンザ流行予測を行うことを目的とする。ピンポイントで予測するためには、気象情報やツイートを用いる研究があるが、本研究において気象情報を用いた予測をした結果はツイートをを用いたものより良い結果は得られなかった。この結果に関しては4.4で示す。そこで、気象情報を用いない予測の一つであるツイートをを用いることで更なる予測精度の向上を目指す。ツイートをを用いた既存研究では対象範囲が広いので、ツイートから地域性を小さい粒度で獲得しているわけではない。そのため、ツイートから地域性を獲得することで、さらなる予測精度の向上を目指す。

ここで、提案手法の概要を図4に示す。Twitterから「インフル」とツイートした人のアカウント名、ツイート、プロフィールを抽出し、アカウント名からユーザータイムラインを取得する。ユーザータイムラインとプロフィールから地域属性判定をする。地域属性を判定したものとインフルエンザ患者数をインプットとし、LSTMネットワークを構築することでインフルエンザ患者の予測をする。

#### 3.2 ツイート処理

Twitterからツイートを収集する処理について示す。ツイートはTwitter社が提供するStreaming APIを用いることで、全ツイートの1%を抽出することができる。また、ツイートともにユーザー名(アカウント名)、ユーザーのプロフィール情報を入手した。そのすべてのツイートに対して「インフル」という単語が含まれるツイートを収集した。インフルエンザの症状には熱、頭痛など様々な症状が現れるが、インフルエンザの流行を表す指標とならないため[20]「インフル」という単語に限定している。この際にリツイートやプライなど直接、患者本人には関係ないツイートを取り除く。

#### 3.3 地域特徴量の抽出

ここではツイート処理をしたものに対して、地域特徴量を抽出した方法を示す。3.2でツイート処理を行ったツイートに対して地域特徴量の抽出を行った。「インフル」というツイートをした人たちが今回の対象範囲調布・府中地域に住んでいたり、足を運んだことがあれば、インフルエンザの感染源になり得る可能性があり、調布・府中地域のインフルエンザ患者数に影響を与えると仮定した。そこで、取得したユーザー名を用いてユーザーごとのタイムラインを取得した。タイムラインの取得にはTwitterのAPIであるtweepyを用いた。タイムライン内に調布・府中地域の情報があれば、仮定した条件に当てはまるとして人数を数えた。調布・府中地域内にある地名や学校、施設の名称の辞書を作り、調布・府中地域の情報が含まれているかどうかを判定することとした。表1に実際に用いた辞書の一部を載せる。「電気通信大学」はツイートの際に「電通大」や「UEC」と略される場合があるので、そういった同じ意味の別単語にも対応した。ツイートだけでなくプロフィール内に地域情報を含んでいる可能性があるためツイートとプロフィール内両方を調べた。プロフィール内には正確な地域情報を含んでいることが多かったが、ツイートには地域名称と同じ名前、例えば「柴崎駅」と「柴崎選手」のように地域名称とそれ以外の固有名詞が含まれているノイズとなりうるツイートがあった。

そこで、地域名称かそうでないかを判断するために、[17]を基とした日本語係り受け解析器、CaboChaを用いた[18]。CaboChaを利用することでツイートの形態素解析、固有表現解析ができる。固有表現解析で「地域」または「施設」と判定されたもののみを地域特徴を有したツイートとみなし数をカウントすることとした。

#### 3.4 LSTM

LSTMをモデル図に表すと図5になる。図5のように、隠れ層にあるLSTMセルの出力を過去の隠れ層として保持し、再度隠れ層に伝播することで再帰的に過去の状態が

表 1: 取得した地名, 施設名称の一部

地域名称	地域名称	施設名称	学校名称
調布	布田	トリエ	電気通信大学
入間町	仙川	味の素スタジアム	東京外語大
上石原	つつじヶ丘	グリーンホール	東京農工大
柴崎	若葉町	深大寺	府中工業

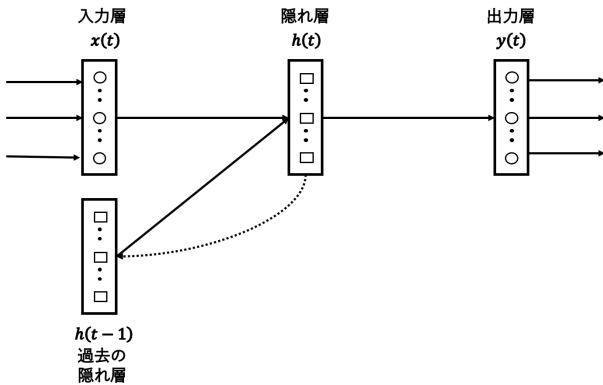


図 5: LSTM の概略図

現在に反映されていることになる。入力にはインフルエンザ患者数と地域特徴を抽出したツイート数を用いる。入力する際にはそれぞれのデータを正規化した。本研究では LSTM ネットワークでは過去の隠れ層を 2 つ前の情報まで用いることにする。つまり、来週のインフルエンザ患者数を予測するのに今週のデータと先週のデータを利用しているということである。これは実験をして精度がよかったのが 2 つ前までの情報を用いることであったからである。

## 4. 実験

### 4.1 実験内容

提案手法と精度を比較するために以下の手法と比較した。

- (1) Influ-Number : インフルエンザ患者数のみを用いて LSTM でモデルを構築したもの
- (2) Weather : インフルエンザ患者数+気象情報を用いて LSTM でモデルを構築したもの
- (3) Tweet-Raw : インフルエンザ患者数+「インフル」を含んだ発言数を用いて LSTM でモデルを構築したもの
- (4) Tweet-Localization : 提案手法

それぞれの手法では、提案手法と同様の方法で LSTM に入力として与えた。ただし、それぞれの手法でのエポック数は異なり、最適なものを利用した。それぞれのエポック数を表 2 に示す。また、隠れ層の数は 300 とした。

本実験では値が正規化されたものが出力されるため、正規化されたものを元に戻し、評価手法に適用した。

### 4.2 データセット

本章では実験で扱うデータについて述べる。本研究では、日本の東京都調布市・府中市周辺に焦点を当てている。

表 2: それぞれの手法のエポック数

	epoch
Influ-Number	5000
Weather	2000
Tweet-Raw	2000
Tweet-Localization	2000

表 3: 実験結果

	MAE	RMSE	$R^2$
Influ-Number	131.7	228.6	0.8879
Weather	135.8	179.7	0.9113
Tweet-Raw	128.9	162.3	0.9354
Tweet-Localization	90.03	144.3	0.9516

ただし、このアプローチはほかの地域にも適用できることができる。データは 2014 年 11 月 10 日から 2015 年 1 月 31 日までと 2017 年 10 月 16 日から 2018 年 6 月 17 日となっている。これはツイートが収集できなかった期間があるためである。使用するデータはツイートデータ、気象データ、インフルエンザ患者数の 3 種類である。

#### 4.2.1 ツイートデータ

Twitter 社からデータを収集した。Twitter の Streaming API を利用した。Streaming API では、Public streams を用いることで Twitter 上を流れる公開データのストリームを収集することができる。Public streams を用いたデータに対し「インフル」というワードが含まれるものを収集した。Tweet-Raw で用いているデータは「インフル」という単語が含まれているものである。LSTM のインプットとして渡すのはツイート数であり、1 週間ごとの合計値とした。

#### 4.2.2 気象データ

気象データは気象庁の府中観測所から収集されたデータを用いた。取得した気象データは平均気温、最低気温、平均風速、日照時間、平均湿度、雨量である。ただし平均湿度は府中観測所では観測されていなかったため、東京観測所のデータを利用した。インフルエンザ患者数は 1 週間ごとのデータのため気象データに関しても 1 週間ごとに平滑化し利用することとする。気象データの組み合わせを試していった結果、平均気温、最低気温、日照時間、雨量を入力したものが一番精度が高かったため、それらを利用した。

#### 4.2.3 インフルエンザ患者数

東京都感染症情報センター [13] から収集されたデータは 1 週間から 2 週間ほど遅れて公表される。今回のデータの範囲はデータセットで示した期間を収集した。データは 1 週間単位の合計量で公表される。

### 4.3 評価手法

提案手法の予測精度は次の 3 つの評価手法を用いて評価する。

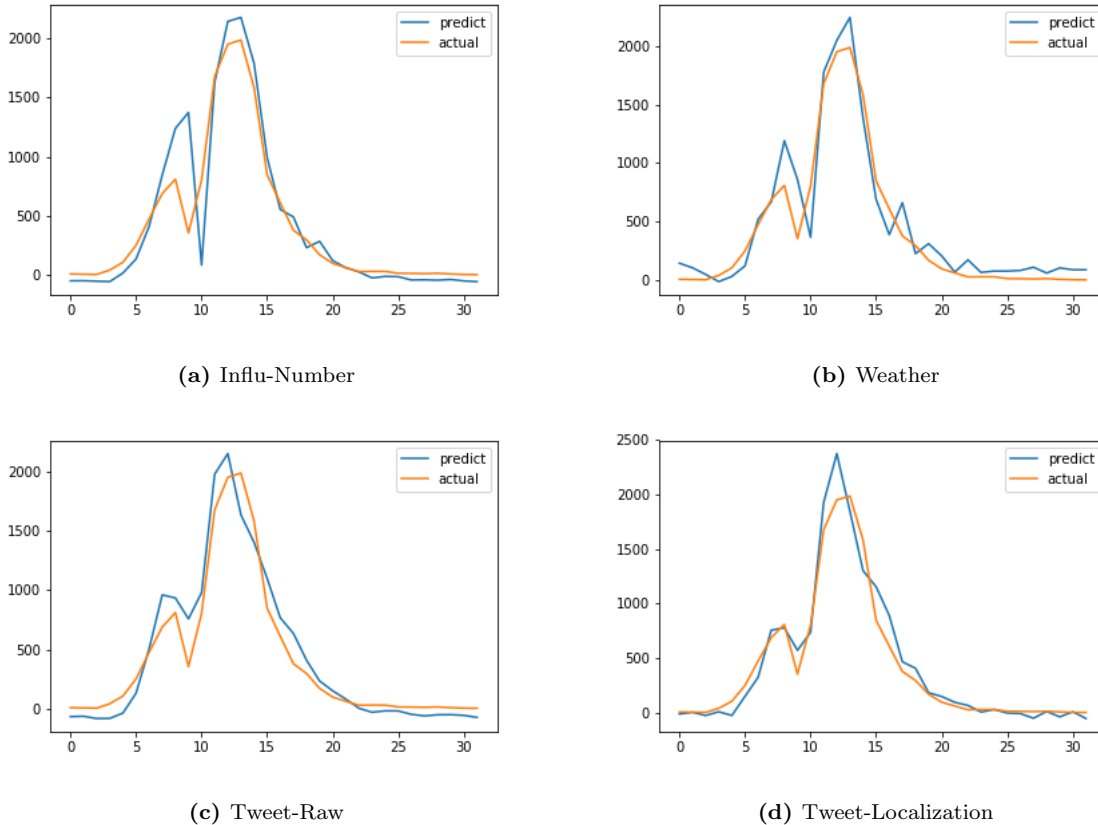


図 6: 実験結果

● 平均絶対誤差 (MAE)

平均絶対誤差 (MAE, Mean Absolute Error) は実測値と予測値の絶対値を平均したものである。MAE が小さいほど誤差が少なく、予測モデルが正確に予測できていることを示す。

$$MAE = \frac{1}{n} \sum_{i=0}^{n-1} |f_i - y_i| \quad (7)$$

● 二乗平均平方根誤差 (RMSE)

二乗平均平方根誤差 (RMSE: Root Mean Squared Error) は実測値と予測値の絶対値の 2 乗を平均したものの (MSE) を平方根で補正したもの。MAE と同じく値が大きいほど誤差の多いモデルであることを示す。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (f_i - y_i)^2} \quad (8)$$

● 決定係数 ( $R^2$ )

決定係数 ( $R^2$ : R-Squared, coefficient of determination) は、モデルの当てはまりの良さを示す指標で、最も当てはまりのよい場合、1.0 となる。

$$R^2 = 1 - \frac{\sum_{i=0}^{n-1} (f_i - y_i)^2}{\sum_{i=0}^{n-1} (\bar{y} - y_i)^2} \quad (9)$$

$n$  はデータ数であり、 $f_i$ ,  $y_i$ ,  $\bar{y}$  はそれぞれ予測値と実測値、実測値の平均を示している。

4.4 実験結果

実験結果を表 3 と図 6 に表す。図は横軸が週を表し、縦軸がインフルエンザ患者数を表す。predict は LSTM で出力された予測した値であり、actual は実際のインフルエンザ患者数を表す。表 3 を見ると、提案手法がすべての評価手法において一番精度が高いことが分かる。Tweet-Raw の手法も精度は気象情報を用いた Weather の手法よりも高く、気象情報よりツイート情報を取り入れた方が精度が上がる事が分かる。

また、図 6 の 5 週目～10 週目の流行初期についても図 6d が実際の流行に遅れることなく正確に予測できている。流行のピークである 12 週目から 14 週目に関してはほかの手法との差は見られなかった。

5. 考察

調布・府中地域において気象情報を用いた予測はあまり精度が高くなかった。つまり、ピンポイント予測する際に、気象情報はあまり役に立たない。これはインフルエンザが寒い冬の時期に流行するという特徴はあるが温度による影響だけではインフルエンザ患者数が正確に表せないことを示唆すると考えられる。また、ウイルスは乾燥時に空気中に浮遊し感染しやすくなるが、今回気象庁から得た平均湿度は 50 パーセントをほとんどの週で超えており、影響が



あまりなかったと考えられる。

Weather と Tweet-Raw を比べると 3 つの指標すべてにおいて Tweet-Raw のほうが精度が高く、気象情報よりツイートのほうが実際のインフルエンザ患者数に影響があると考えられる。ここで、地域判定していない Tweet-Raw がなぜ精度が高かったかを考えると、インフルエンザは場所によって 1 週間や 2 週間ほどピークの位置、初動がずれたりするが全体としてのインフルエンザ傾向は日本において大きな差はないので、精度が高く予想できたのだと考えられる。

Tweet-Raw と提案手法を比べると提案手法のほうが精度が高く、ピンポイントのインフルエンザ予測には地域名を考慮すると精度が高くなるのが分かる。つまり、全体としてのインフルエンザ傾向は確かに一致しているが各地域の一週間単位を見てみると、多少ピークの位置や初動に違いがあるため、地域ごとにあった予測が必要であると考えられる。だから、地域属性を特定することがピンポイントインフルエンザ流行予測に有用である。

## 6. おわりに

### 6.1 本論文のまとめ

本論文では、調布・府中地域という特定の地域においてのインフルエンザ流行予測をするために、ツイートに着目し、予測精度の向上を目指した。地域名称や施設名称でユーザーの地域属性を判定することで、ツイート情報をそのまま用いることよりも予測精度が高くなるのが分かった。つまり、特定の地域の流行予測をする際に地域属性を判定することがピンポイントインフルエンザ流行予測に有用であることが分かった。

### 6.2 今後の課題

今後の課題として、次の 2 点を考えている。1 点目はデータ量の追加である。本研究では、ツイートが集められた期間が約 2 年分であったため、学習データに偏りが不足があった可能性がある。今後はさらに、ツイートを収集していくことでモデルの改善、精度の改善をしていきたい。2 点目は、地域属性の特定の手法である。本研究では、地域属性の特定の精度を高めることでインフルエンザ流行予測の精度が高まると考えられる。地域属性を特定するためにユーザーのツイートだけを対象として実験を行っていて、ユーザーのツイートした時間を考慮できていない。[19] ではツイートだけではなく、フォローフォロワー関係など様々な情報を取り入れることで地域特定をすることで地域属性の精度を高めることができる。これを利用することでユーザーの現在位置など時間を考慮した分析をすることができ予測精度の向上につながると考えられる。

謝辞 本研究は JSPS 科研費 JP17H04705, JP18H03229, JP18H03340, JP18K19835, JP19H04113, JP19K12107 の

助成を受けたものです。

## 参考文献

- [1] 厚生労働省:平成 30 年度 今冬のインフルエンザ総合対策について入手先 (<https://www.mhlw.go.jp/bunya/kenkou/influenza/index.html>) (2019/09/20 参照)
- [2] 厚生労働省:新型インフルエンザに関する Q&A 入手先 (<https://www.mhlw.go.jp/bunya/kenkou/kekkaku-kansenshou04/02.html>) (2019/09/20 参照)
- [3] ARAMAKI, Eiji; MASKAWA, Sachiko; MORITA, Mizuki. Twitter catches the flu: detecting influenza epidemics using Twitter. In: Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics, 2011. p. 1568-1576.
- [4] ALBAYRAK, Maya D.; GRAY-RONCAL, William. Data Mining and Sentiment Analysis of Real-Time Twitter Messages for Monitoring and Predicting Events. In: 2019 IEEE Integrated STEM Education Conference (ISEC). IEEE, 2019. p. 42-43.
- [5] 橋本康弘, et al. 都市におけるジオタグ付きツイートの統計 (<特集> 人と環境に見る高次元データフローの生成と解析). 人工知能学会誌, 2012, 27.4: 424-431.
- [6] S. R. Venna, A. Tavanaei, R. N. Gottumukkala, V. V. Raghavan, A. S. Maida and S. Nichols, "A Novel Data-Driven Model for Real-Time Influenza Forecasting," in IEEE Access, vol. 7, pp. 7691-7701, 2019.
- [7] 東京都インフルエンザ情報 第 22 巻 第 1 号入手先 ([idsc.tokyo-eiken.go.jp/assets/flu/2019/Vol22No1.pdf](http://idsc.tokyo-eiken.go.jp/assets/flu/2019/Vol22No1.pdf)) (2020/01/10 参照)
- [8] H. W. Hethcote, "The mathematics of infectious diseases", SIAM Rev. vol. 42, no. 4, pp. 599-653, 2000.
- [9] M. B. Hooten, J. Anderson, L. A. Waller, "Assessing North American influenza dynamics with a statistical sirs model", Spatial Spatio-Temporal Epidemiol., vol. 1, no. 2, pp. 177-185, 2010.
- [10] G. Chowell, M. A. Miller, C. Viboud, "Seasonal influenza in the united states France and Australia: Transmission and prospects for control", Epidemiol. Infection, vol. 136, no. 6, pp. 852-864, 2008.
- [11] SOEBIYANTO, Radina P.; ADIMI, Farida; KIANG, Richard K. Modeling and predicting seasonal influenza transmission in warm regions using climatological parameters. PloS one, 2010, 5.3: e9450.
- [12] A. F. Dugas et al., "Influenza forecasting with Google flu trends", PLoS ONE, vol. 8, no. 2, pp. e56176, 2013.
- [13] 東京都感染症情報センター: 東京都感染症発生動向調査, 入手先 (<https://survey.tokyo-eiken.go.jp/epidinfo/epimenu.do>) (2019/09/20 参照)
- [14] 気象庁:過去の気象データ検索, 入手先 (<http://www.data.jma.go.jp/obd/stats/etrn/index.php>) (2019/09/20 参照)
- [15] S. Hochreiter, and J.Schmidhuber. Long short-term memory. Neural Computation, 9(8), pp.1735-1780, 1997.
- [16] F. A. Gers, J. Schmidhuber. and F. Cummins. Learning to forget: continual prediction with LSTM. Neural Computation, 12(10), pp.2451-2471, 2000.
- [17] KUDO, Taku; MATSUMOTO, Yuji. Fast methods for kernel-based text analysis. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1. Association for Computational Linguistics, 2003. p. 24-31.
- [18] CaboCha/南瓜: Yet Another Japanese

Dependency Structure Analyzer 入手先  
(<https://taku910.github.io/cabocho/>) (2020/01/16  
参照)

- [19] ZHENG, Xin; HAN, Jialong; SUN, Aixin. A survey of location prediction on twitter. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30.9: 1652-1671.
- [20] 北村有里恵, et al. 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. *Ann. Rep. Tokyo Metr. Inst. Pub. Health*, 2017, 68: 66.