

アメダスの観測データを用いた Twitter ユーザの居住地推定の試み

近藤 佑樹^{†,a} 萩行 正嗣^{†,b} 吉田 光男^{†,c} 梅村 恭司^{†,d}

[†]豊橋技術科学大学 [‡]株式会社ウェザーニューズ

a) y153331@edu.tut.ac.jp b) hangyo@wni.com c) yoshida@cs.tut.ac.jp d) umemura@tut.jp

概要 ソーシャルメディアの投稿を活用する際に、ユーザの居住地と投稿内容とを関連付けることにより、より有用な情報を得ることができる。しかし自身の居住地を登録しているユーザは少なく、ユーザの居住地を推定する研究が広く行われている。本研究では、アメダスによる日本全国の観測データを利用して Twitter ユーザの居住地を推定する手法を提案する。本手法ではまず、ツイートの投稿テキストから、投稿ユーザがどのような天気の中で投稿したかを推定する。次に推定した天気および投稿日時を各地のアメダスによる観測データと照合する。最後にどの地域から集中的に投稿しているかを推定し、その地域をユーザの居住地とみなす。単語の地理的局所性を利用する従来の手法と比較し、良好な結果が得られた。

キーワード Twitter, 位置情報推定, アメダス, 気象情報

1 はじめに

Twitter¹ は、短文（ツイート）を投稿できるソーシャルメディアである。投稿されるツイートの数が多く、現実世界のイベントを反映したソーシャルセンサとみなすことができる [1, 2]。位置情報と関連付けられたソーシャルセンサとして利用するために、ユーザの居住地を推定する試みが広く行われている。

ユーザの居住地を推定する手法として、ユーザが投稿したツイートを用いるものが提案されている [3, 4]。このような手法の多くは、ユーザが投稿したツイートに含まれる単語のうち、出現する地域（エリア）に偏りのあるものを手掛かりに居住地を推定する。ツイートに含まれる単語の多くが特定のエリアでのみ出現する場合、ユーザの居住地はそのエリアである可能性が高いと推定できる。しかし、出現するエリアに偏りのある単語を含むツイートを投稿していないユーザについては、居住地を推定することが困難である。

本稿では、出現するエリアに偏りがある単語を含まないものの、天気に関する情報を含んでいるツイートに着目する。例えば「雨降ってる」というツイートの場合、その投稿日時において、ユーザがいた地点では雨が降っていたと推測できる。このようなツイートを利用し、ユーザが投稿したツイートに対して、それらが投稿された状況における天気を推定し、推定した天気の情報と投稿日時からユーザの居住地を推定する手法を提案する。また、出現するエリアに偏りのある単語を手がかりとする手法と比較し、都市部以外のユーザの居住地を推定する場合において、提案手法が有効に機能したことを示す。

2 提案手法

2.1 概要

本手法では、まず、ツイートの投稿内容とツイートが投稿された地点・日時における天気との関係を次の手順で学習する。まず、位置情報付きツイートに対して分かち書きを行い、特徴ベクトルに変換する。次に、アメダスの観測データを用いて、ツイートが投稿されたエリアでの投稿された日時における天気をラベルとして付与する。最後に、作成したラベル付き特徴ベクトルをサポートベクターマシン (SVM) に学習させる。

次に、居住地推定の対象であるユーザが投稿したツイートを収集し、それらに対して SVM を用いてラベルを推定する。このラベルは、推定対象ユーザがツイートを投稿したエリア・日時における天気を表している。この天気の情報在全国のアメダスにおける観測データと比較し、同様の天気が観測されたエリアをユーザの居住地と推定する。

2.2 ツイート内容と天気の関係の学習

まず、分かち書きをしたツイート t を特徴ベクトル \mathbf{x}_t に変換する。

$$\mathbf{x}_t = (x_1, x_2, \dots, x_{|W|})$$

$$x_i = \begin{cases} 1 & w_i \in W_t \\ 0 & (\text{otherwise}) \end{cases}$$

ここで W は学習に用いたツイート全てを通して 2 回以上出現した単語の集合、 W_t はツイート t に含まれる単語の集合である。また w_i は W に含まれる単語を辞書順に列挙した場合に i 番目となる単語である。

次に、作成した特徴ベクトル \mathbf{x}_t に対して、アメダスの観測データを参照してツイートが投稿されたエリア・

Copyright is held by the author(s).

The article has been published without reviewing.

¹<https://twitter.com/> (accessed 2017-4-20)

日時における天気をラベルとして付与する。ラベル $l = \{\text{True}, \text{False}\}$ であり、 $l = \text{True}$ は投稿されたエリアでその時間に雨が降っていたことを、 $l = \text{False}$ は雨が降っていなかったことを表す。

最後に、作成したラベル付き特徴ベクトルを SVM に学習させる。本稿では、線形カーネルを利用し、コストパラメータを $C = 0.025$ とした。

2.3 天気の情報を利用した居住地の推定

天気の情報を用いてユーザの居住地を推定するには、まず、推定対象となるユーザが投稿したツイートを収集し、それらが投稿された地点・日時における天気を推定する。その後、推定した天気の情報を用いて居住地を推定する。

居住地推定の対象となるユーザが投稿したツイートの集合を $T = \{t_1, t_2, \dots, t_n\}$ とし、ツイート $t_j \in T$ が投稿された日時を $date(t_j)$ とする。また、ツイート t_j が投稿されたエリア・日時における推定した天気を $w\text{-est}(t_j)$ とする。天気の推定では、ツイート t_j から生成した特徴ベクトル x_{t_j} に対してラベル $l \in \{\text{True}, \text{False}\}$ を推定する。つまり $w\text{-est}(t_j) \in \{\text{True}, \text{False}\}$ である。

エリアの集合を $A = \{a_1, a_2, \dots, a_m\}$ とし、ユーザの居住地は A の要素のうちいずれか 1 つであるとする。ここで m は全エリア数であり、各エリアはアメダスが設置されている地点と 1 対 1 に対応する。アメダスの観測データを参照して得られた、エリア a における日時 k での天気を $w\text{-obs}(a, k)$ とすると、 $w\text{-obs}(a, k) \in \{\text{True}, \text{False}\}$ となる。このとき、ユーザの居住地 a_{est} を次式で推定する。

$$a_{est} = \arg \min_{a \in A} count(T, a)$$

$$count(T, a) = \sum_{t \in T} diff(w\text{-est}(t), w\text{-obs}(a, date(t)))$$

$$diff(l_1, l_2) = \begin{cases} 0 & l_1 = l_2 \\ 1 & l_1 \neq l_2 \end{cases}$$

3 データの準備

3.1 アメダスの観測データ

アメダス (AMeDAS) とは、気象庁が運営している、地域の気象を自動で観測するシステムである。日本全国で約 1,300 か所に設置されている観測地点での降水量や気温、風向・風速などの観測を行っており、観測データは気象庁のサイトで公開されている。しかし、降水量の観測データは 0.5mm 単位で記録されており、0.5mm 未満の雨と降水なしとが区別できない。

本稿では、感雨計による詳細な降水量の観測データを利用し、雨が降っていたかどうかのラベル付けを行う。感雨計とは、0.5mm 未満の僅かな降水の有無を検知す

る設備であり、全国 157 か所の気象官署に設置されている。これにより、全国のアメダスによる観測データのうち、気象官署における降水量の観測データでは、0.5mm 未満の降水と降水なしを区別できるようになっている。

3.2 エリアとツイートデータ

地上気象観測地点一覧²に記載されている気象官署 157 か所のうち、昭和基地、富士山、阿蘇山の 3 地点を除いた 154 か所から半径 10km の範囲をエリアとしている。ツイートの投稿位置とユーザの居住地は、共にこれらのエリアのうちいずれかであるとする。各エリアの中心の緯度、経度などのデータは、気象庁のサイト³で公開されているものを用いる。2016 年に投稿された位置情報付きツイートに対して、次の処理を施す。

1. GPS により付与される、緯度・経度の座標からなる位置情報を持たない (“coordinates” に値が無い) ツイートを除去
2. クライアント名 (“source”) により、Bot または特定のサービスにより投稿された定型文であると判断したツイートを除去
3. 緯度・経度の値から、半径 10km 以内にある最寄りのアメダスを探し、その ID をエリア ID として付与 (エリア ID を付与できない場合はそのツイートを除去)
4. エリア ID と投稿日時から、雨が降っていたかどうかのラベルを付与 (欠損値により付与できない場合はそのツイートを除去)
5. 投稿内容から、ハッシュタグ、URL、ユーザ ID を除去し、これにより空文字列または空白文字のみとなったツイートを除去
6. 投稿内容を MeCab+IPA 辞書で分かち書きし、単語のリストに変換
7. 後述する 3.3 節に示す、天気に関連する単語を 1 つも含まないツイートを除去

以上の処理を行ったのち、残ったツイートをユーザごとに集計した。集計の結果、10 ツイート以上残っているユーザに対して、9 割以上のツイートが同じエリアからの投稿である場合、そのエリアをユーザの居住地とした。居住地を付与できないユーザとそのツイートは除いた。

本稿では、2016 年に投稿された位置情報付きツイート 171,614,139 件のうち、153 ユーザが投稿した 4,613 ツイートを実験に使用する。

²<http://www.data.jma.go.jp/obd/stats/data/mdrr/chiten/sindex2.html> (accessed 2017-4-20)

³<http://www.jma.go.jp/jma/kishou/known/amedas/kaisetsu.html> (accessed 2017-4-20)

3.3 天気に関連する単語の収集

提案手法では、推定対象ユーザが投稿したツイートについて、それが投稿された地点・日時における天気を推定し、推定した天気を元にユーザの居住地を推定する。しかし、天気の情報を含んでいないツイートについては天気を推定することができない。本稿では、実験の前に、天気の情報を含まないツイートを除去する。

天気の情報を含まないツイートを除去するため、日常会話において天気を表すために用いられる単語に注目した。私たちは、日常会話において「天気が晴れである」ことについて言及するときに「晴れ」「良い天気」などの言葉を用いる。したがって、このような天気を表すために用いられる単語が含まれているツイートは、投稿された地点と日時における天気の情報を含んでいると考えられる。本稿では、天気を表すために用いられる単語をツイート中から収集し、収集した単語を含むツイートのみを用いる。

天気を表すために用いられる単語は、天気が晴れまたは雨である日に偏って出現すると考えられる。よって、天気を表すために用いられる単語は、単語が出現したツイートに付与されているラベルが True と False のどちらかに偏っていると考えられる。

W_t をツイート t に含まれる単語の集合とする。ラベル l はツイート t に対して付与されるが、これを W_t に含まれる単語すべてに対してラベル l が付与されているとみなす。各単語について以下の2つの事象の自己相互情報量 (PMI) を計算し、単語の偏り度合いを評価する。

- その単語が出現した
- 単語のラベルが True (または False) である

単語 w に対するラベル l の付きやすさ $\text{PMI}(w, l)$ の定義を以下に示す。

$$\begin{aligned} \text{PMI}(w, l) &= \log \frac{P(w, l)}{P(w)P(l)} \\ P(w, l) &= \frac{C(w, l)}{N} \\ P(w) &= \frac{C(w)}{N} \\ P(l) &= \frac{C(l)}{N} \end{aligned}$$

ここで $C(w, l)$ は単語 w のラベルが l であった回数、 $C(w)$ は単語 w の総出現回数、 $C(l)$ はラベル l の総出現回数、 N は全ての単語の出現回数の総和である。

前述した 3.2 節の処理 5 まで実施したツイート集合において、ツイート中に 10 回以上出現している単語について、 $l = \text{True}$ と $l = \text{False}$ の場合について $\text{PMI}(w, l)$ を計算し、値の大きい順に 2,000 単語ずつを集めた。し

表 1 ラベルが F であるツイートに多く出現する単語

出現回数	ユーザ数	PMI	単語
33	30	0.6133	あつい
32	21	0.6133	晴天
10	10	0.6784	炎天下
13	11	0.6784	暁
405	142	0.6867	U+1F338
26	22	0.6922	眩しい
126	90	0.7235	暑く
14	14	0.7780	映える
21	19	0.7780	初日の出
86	71	0.7780	日和
25	15	0.8077	U+26C5
39	32	0.8127	日差し
75	57	0.8159	快晴
11	10	0.8628	日比谷公園
20	12	0.8934	U+1F349
15	14	0.9008	絶好
10	10	1.0003	日陰
15	12	1.0003	朝晩

※ U+1F338, U+26C5, U+1F349 はそれぞれ cherry blossom, sun behind cloud, watermelon と呼ばれる絵文字である。

しかし、出現頻度の小さい単語が多く混ざるといった問題があった。これは、出現頻度が小さい単語には、天気に関係なく出現するにも関わらず $P(w, \text{True})$ と $P(w, \text{False})$ に大きな差が生じ、PMI が過剰に高くなる傾向があるためである。そこで、各単語についてその単語を一度でも使ったユーザの数が 10 未満である単語を除去することとした。残った単語の中から、ラベルが True であるツイートに偏って出現しやすい単語 50 種類と、ラベルが False であるツイートに偏って出現しやすい単語 18 種類を手動で選んだ。選んだ単語とその出現回数、単語を使用したユーザ数、PMI を表 1 と表 2 に示す。

4 実験設定・結果

4.1 ベースライン手法

提案手法の有効性を確認するため、エリアごとの単語の分布を用いる居住地推定手法と比較する。異なる観点で比較するために、2種類の手法をベースラインとして性能を比較する。

単語のエリアごとの出現確率を利用する手法として、Cheng らが提案した手法がある [3]。これは、ユーザが投稿したツイートに含まれる単語について各エリアにおける出現確率を推定し、単語の出現確率の和が最大となるエリアをユーザの居住地と推定する手法である。この手法をベースライン (A) とする。

提案手法は、まずユーザが投稿したツイートに対して天気を推定し、推定した天気を元に居住地を推定するという 2 段階の手法である。そこで、提案手法と同様の、2 段階に分かれた推定手法とも性能の比較を行う。森國らは、Cheng らの手法を元にしたツイートの投稿位置を推定する手法を提案している [4]。森國らの手法を元に

表 2 ラベルが T であるツイートに多く出現する単語

出現回数	ユーザ数	PMI	単語
123	98	0.7411	濡れ
106	50	0.7444	雪かき
75	50	0.7481	積雪
138	101	0.7511	積もっ
13	11	0.7586	降り
26	24	0.7586	長靴
15	12	0.7652	雨天
28	26	0.7773	びしょ濡れ
207	151	0.7850	大雨
70	56	0.7873	豪雨
111	81	0.7902	大雪
30	27	0.7932	ずぶ濡れ
38	22	0.7932	U+1F302
26	13	0.8152	U+2602
4019	1798	0.8171	雨
1150	720	0.8198	降り
26	25	0.8228	降り出し
18	16	0.8297	夕立
33	28	0.8297	止ま
192	128	0.8344	雷
47	39	0.8410	吹雪
28	24	0.8421	吹雪い
528	199	0.8469	U+2614
10	10	0.8476	ちらつい
20	20	0.8476	雨風
11	11	0.8621	降り続く
47	46	0.8741	雨宿り
13	10	0.8842	増水
39	27	0.8842	雪だるま
52	49	0.8842	土砂降り
38	22	0.8872	パラパラ
14	12	0.8927	しとしと
14	12	0.8927	ザーザー
13	12	0.8927	風雨
116	76	0.8956	あいにく
30	23	0.9001	積もる
33	23	0.9095	休校
178	96	0.9097	小雨
50	40	0.9104	生憎
17	15	0.9122	雪合戦
86	59	0.9132	雨降り
35	26	0.9148	本降り
57	40	0.9216	どしゃ降り
41	34	0.9275	止む
21	16	0.9325	やむ
22	16	0.9325	暴風雨
28	13	0.9408	U+1F301
14	12	0.9472	暴風
10	10	0.9997	びしょびしょ
13	12	0.9997	rain

* U+1F302, U+2602, U+2614, U+1F301 はそれぞれ closed umbrella, umbrella, umbrella with rain drops, foggy と呼ばれる絵文字である。

した居住地推定手法をベースライン (B) とし、提案手法と推定性能を比較する。具体的には、まずユーザが投稿した各ツイートについて、森國らの手法を用いて投稿位置を推定する。その後、推定されたエリアの中から多数決で選ばれたエリアをユーザの居住地と推定するものである。

なお、双方のベースライン手法に、単語のフィルタリングおよびスムージングを適用することが出来るものの、本実験ではそれらを適用せずに居住地を推定する。

4.2 評価指標

推定性能の評価には適合率 Precision@k を用いる。これは、各ユーザについて居住地を推定したとき、推定したエリアの上位 k 個に実際の居住地が含まれているユーザの割合である。ユーザの推定した居住地が、実際の居住地であるかの評価には、正解距離を用いる。これは、推定した居住地と実際の居住地との距離が正解距離未満であれば正しく推定できたとみなすものである。ユーザの集合 U に対して正解距離 d で居住地を推定したときの Precision@k(U, d) は以下の式で表される。

$$\text{Precision@k}(U, d) = \frac{|\{u|u \in U \wedge A_t(u, d) \neq \phi\}|/|U|}{A_t(u, d) = |\{a|a \in A_k(u) \wedge \text{dist}(a, a_t(u)) < d\}|}$$

ここで $a_t(u)$ はユーザ u の実際の居住地、 $A_k(u)$ はユーザ u に対して推定した居住地のうち上位 k 個からなる集合である。また $\text{dist}(a_1, a_2)$ はエリア a_1, a_2 間の距離である。

都道府県ごとの推定性能の評価も同様に Precision@k を用いる。ただし、ユーザの集合 U をユーザの実際の居住地が含まれる都道府県ごとに分類し、都道府県ごとに適合率を評価する点が異なる。都道府県 p が実際の居住地であるユーザの集合 U_p は以下で表される。

$$U_p = \{u|u \in U \wedge \text{pref}(a_t(u)) = p\}$$

ただし P は日本にある都道府県の集合、 $\text{pref}(a)$ は a が含まれる都道府県である。全ての $p \in P$ について Precision@k'(U_p, d) を以下の式で算出する。

$$\text{Precision@k}'(U_p, d) = \frac{|\{u|u \in U_p \wedge A'_t(u, d) \neq \phi\}|/|U|}{A'_t(u, d) = |\{a|a \in A_k(u) \wedge \text{dist}(a, a_t(u)) < d \wedge \text{pref}(a) = \text{pref}(a_t(u))\}|}$$

4.3 正解距離と適合率との関係

$k = 1, 3, 5$ とし、正解距離 d を 10km から 160km まで変えたときの適合率 (Precision@k) をそれぞれ図 1, 図 2, 図 3 に示す。

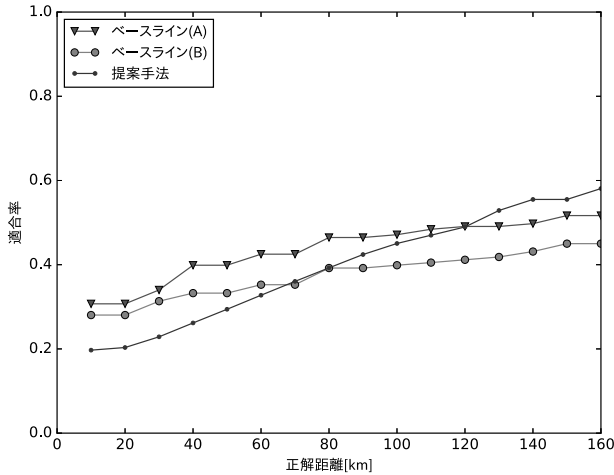


図1 正解距離を変えたときの適合率 ($k = 1$)

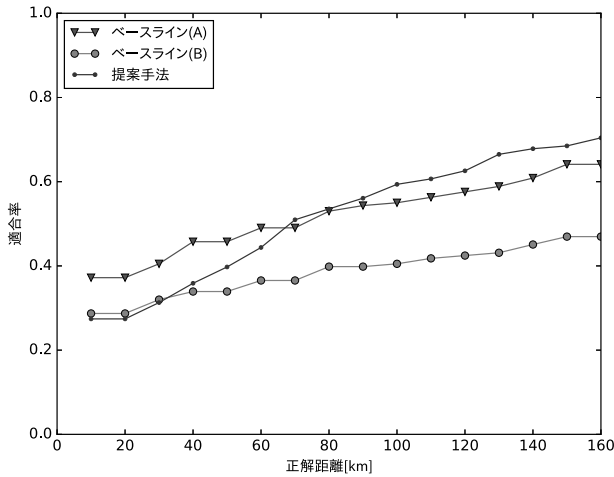


図2 正解距離を変えたときの適合率 ($k = 3$)

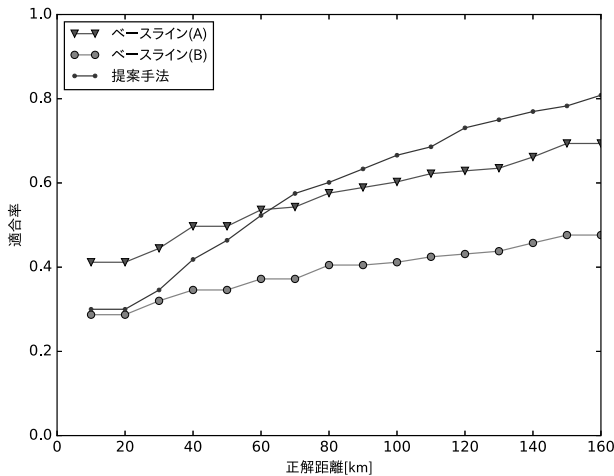


図3 正解距離を変えたときの適合率 ($k = 5$)

$k = 1, 3, 5$ のすべての場合において、ベースライン (A) がベースライン (B) を上回っている。ベースライン (B) は、各ツイートについて投稿された位置を推定し、推定した結果から多数決で居住地を推定する手法である。ツイートの投稿位置を誤って推定したことが適合率の低下につながったと考えられる。

提案手法は、正解距離がおおよそ 60km から 80km より大きいときにベースライン (A) の適合率を上回っている。また、ベースラインと比べてグラフの傾きが急であり、正解距離を大きくしたときの適合率の向上がより顕著である。これは、正解距離を大きくしたときに、居住地の推定結果が不正解から正解へと変化したユーザの数の差を表しており、ベースラインではこのようなユーザが少なく、逆に提案手法では多いといえる。つまり提案手法においては、例え不正解であったとしても、ベースラインに比べ、実際の居住地に近いエリアを推定していることが示唆される。

4.4 都道府県別の適合率

$k = 1$, 正解距離 $d = 10$ として都道府県別に評価した適合率 (Precision@ k') を図 4 に示す。ここで、その都道府県に実際の居住地が含まれるようなユーザの数が 0 であった都道府県は表示していない。また県名は 2016 年度の都道府県別人口⁴を参考に、人口が多い順にソートしている。この結果から、ベースラインは (A)(B) ともに人口が多い都道府県でしか正解できていないことがわかる。また提案手法では、ベースラインに比べてより多くの都道府県で正解できている。このことは、マクロ平均からも明らかであり、提案手法が 0.242、ベースライン (A) が 0.0689、ベースライン (B) が 0.0458 であった。

ユーザの居住地を推定する手法の多くは、エリアごとの単語の出現頻度を用いて、単語の出現確率を推定している。単語のうち、一部のエリアでの出現確率が高い単語はローカルワードと呼ばれる。コンテンツベースの居住地推定手法の多くはローカルワードを手掛かりとして居住地を推定している。ところが、位置情報付きツイートは約 40%が東京近郊で投稿されており、ツイートは人口以上に大都市、特に東京近郊に集中している [5]。エリアによって投稿されるツイート数が異なるため、実際の出現確率がどのエリアでも等しいような単語が、人口の多いエリアでの出現確率が高く推定される。したがって、ローカルワードを含んだツイートを投稿していないユーザは、投稿されるツイート数が多いエリアが居住地であると推定される。

実験結果から、ベースラインでは多くのユーザについて

⁴https://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&listID=000001154737&requestSender=estat (accessed 2017-5-15)

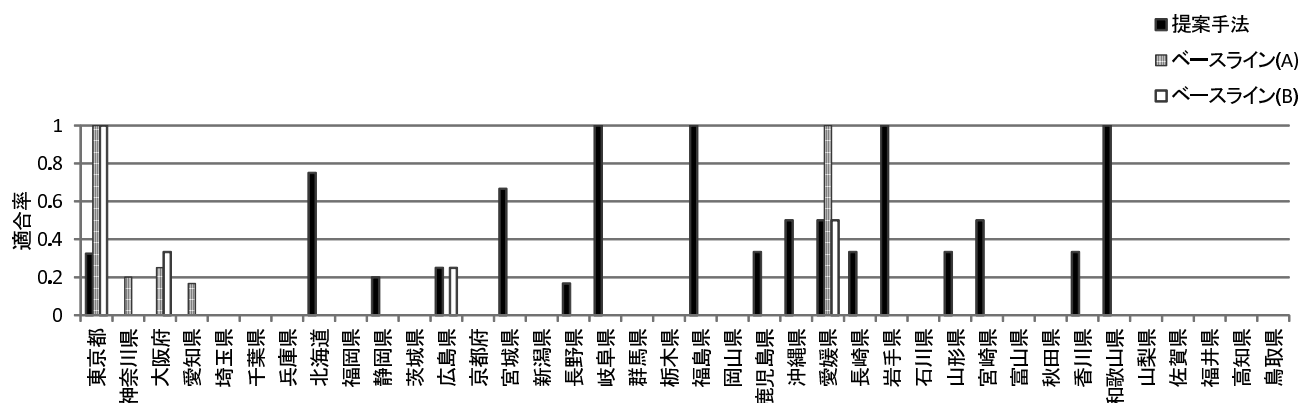


図4 都道府県別の適合率

て人口の多いエリアが居住地であると推定されたことがわかる。実験には天気に関する単語を含むツイートのみを用いており、今回、ユーザはローカルワードを含むツイートをほとんど投稿しないと見なせる実験設定になっている。以上より、ローカルワードを含まず天気に関する単語を含むようなツイートを多く投稿しているユーザに対しては、ベースラインより提案手法の方が有効である。

5 関連研究

Twitter をソーシャルセンサとみなし、イベントや流行を検知するための研究開発が盛んに行われている [1, 2]. Sakaki らは、日本で発生した震度 3 以上の地震の震源地を高精度に特定している [6]. Aramaki らは、インフルエンザの流行を検出する方法を提案し、流行状況を可視化している⁵[7].

ソーシャルセンサとして利用するためには、ユーザの居住地が必要になるが、それらの情報があらかじめ付与されているケースは少ない [3]. そのため、ツイートなどの情報を利用し、ユーザの居住地を推定する試みがなされている [8]. このような手法の多くは、ユーザが投稿したツイートに含まれる単語のうち、出現するエリアに偏りのあるものを手掛かりに居住地を推定する。

従来の手法では、出現するエリアに偏りのある単語を手がかりにするものの、そもそも、その情報を学習するための居住地の情報が不足している傾向があった。本稿では、天気の情報という、あらかじめ場所を特定することのできるイベントを利用することで、居住地を推定する手法を提案した。

6 おわりに

本稿では、エリアごとの単語の分布を使用しない居住地推定の手法として、ツイートに関する天気を推定し、その推定した天気を用いる手法を提案した。

提案手法と単語の分布を使用した既存手法とで推定性能を比較したところ、推定エリア上位 2 件以上、エラー距離 70km 以上の条件下で、提案手法が従来手法を上回る精度で居住地を推定できた。さらに、都道府県ごとの推定性能を比較したところ、既存手法では人口の多い一部の都道府県でしか正解できていないのに対して、提案手法では人口の少ない都道府県でも正解できているという結果を得た。これらの実験結果から、既存手法では居住地推定が行えないが、提案手法により推定が行える場合があることが分かった。

参考文献

- [1] 榎剛史, 松尾豊: ソーシャルセンサとしての Twitter - ソーシャルセンサは物理センサを凌駕するか?-, 人工知能学会誌, Vol. 27, No. 1, pp. 67-74, 2012.
- [2] Atefeh, F. and Khreich, W.: A Survey of Techniques for Event Detection in Twitter, Computational Intelligence, Vol. 31, No. 1, pp. 132-164, 2015.
- [3] Cheng, Z., Caverlee, J., and Lee, K.: You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users, Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 759-768, 2010.
- [4] 森國泰平, 吉田光男, 岡部正幸, 梅村恭司: ツイート投稿位置推定のための単語フィルタリング手法, 情報処理学会論文誌: データベース, Vol. 8, No. 4, pp. 16-26, 2015.
- [5] 橋本康弘, 岡瑞起: 都市におけるジオタグ付きツイートの統計, 人工知能学会誌, Vol. 27, No. 4, pp. 424-431, 2012.
- [6] Sakaki, T., Okazaki, M. and Matsuo, Y.: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, Proceedings of the 19th International Conference on World Wide Web, pp. 851-860, 2010.
- [7] Aramaki, E., Maskawa, S. and Morita, M.: Twitter Catches The Flu: Detecting Influenza Epidemics using Twitter, Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pp. 1568-1576, 2011.
- [8] Zheng, X., Han, J. and Sun, A.: A Survey of Location Prediction on Twitter, arXiv.org, arXiv:1705.03172, 2017.

⁵http://mednlp.jp/influ_map/ (accessed 2017-6-9)