

ソーシャルメディアにおける公共アカウントに着目したユーザの場所分析

Analysis of User Location in Social Media Focused on Public Services Account

榎田 宗丈[†] 吉野 孝[†]
Sojo Enokida Takashi Yoshino

1. はじめに

2011年3月11日に発生した東日本大震災ではTwitter^{*1}が情報を得るためのメディアとして活用され[1], 新戦略推進専門調査会(内閣府)の防災・減災分科会ではTwitterをはじめとするSNS等を防災・減災に有効活用する課題や目的が報告されている。Twitterが防災・減災に活用された事例を踏まえて, 多くの自治体や団体が災害情報等を発信するTwitterアカウントの開設を行い, Twitterを使った自治体による訓練などが行われている^{*2}。2016年4月14日に発生した平成28年(2016年)熊本地震(以下, 熊本地震)では, Twitter Japanの調査によって熊本地震に関連するツイート数が東日本大震災直後のツイート数よりも上回っていることが分かっている^{*3}。このように, Twitterを災害時および防災において, 情報を発信するメディアとして利用することが増えている。

Twitterの地域住民の発信した情報は, 災害時の重要な情報源となる。災害時に地域住民が発信した情報は, 被災地域の地域住民や災害担当者が被害状況を把握する, 救助や安否確認に利用する, 他の地域の人々が被災地の状況を知り支援を行う, などとして活用される。そのため, ユーザの場所を事前に推定することで, 災害時に地域の情報をより迅速に収集することができるようになる。

Twitterユーザの場所を推定する方法としては, ツイートに付与されているジオタグを用いる方法, ユーザのプロフィールを用いる方法, ツイート本文に含まれる地名を取得する方法, ソーシャルグラフを利用する方法などがある[2]。本稿ではRobらのSocial Triangulation[3]をもとに, ユーザの位置の推定を行う。Social Triangulationは, 地域住民は自分の住んでいる地域の公共サービスなどの地域組織のアカウントをフォローしていると仮定し, ユーザの位置を推定する方法である。Robらは, 調査した特定の市の地域組織をフォローしているユーザの中で, プロフィールに場所情報を登録しているユーザ79,998人のうち, 68%を地域住民として特定できたとしている。

本稿では, Twitter ライフライン (@TwitterLifeline)^{*4} のリストでまとめられている各都道府県の災害時に役に立つアカウント一覧から, 各都道府県の公共アカウントを取得し, その公共アカウントのフォロワーがその地域のユーザであることを確認する。ユーザの地域を確認する方法としては, ユー

ザがプロフィールに登録している場所情報から, JUMAN++で形態素解析をして, 地名と分類される形態素があるかどうかを確認をしている。本稿の目的は, 災害が多く, 災害時の情報発信として日本で特に活用されるTwitterにおいて, 公共アカウントをフォローしていることによって, ユーザの位置推定ができるかを検証することである。

2. 関連研究

宮部ら[4]らは, 東日本大震災のツイート地域ごとに分類して, リツイート分析を行っている。これより, 被災地から多くの情報が発信され, 特に被害の大きい地域で発信された情報については, 他地域へと移動する傾向が見られている。

Sonら[5]は, 東日本大震災時には地震, 津波および原発事故とツイート数に高い相互関係があることを示しており, Twitterが災害時の早期警告に有用な資源を持つだけでなく, 社会不安やニーズを分析するツールとなりえるとしている。

小川ら[6]は, Twitterと地域SNSの比較を行っており, Twitterのユーザは地震情報, 食品・物資, 放射能などといったどのユーザにも共通する話題を共有するとしている。また, 被災地域, 非被災地域別の話題について比較しており, 直接的な被害のあった被災地域であるほど安否・無事・心配に関する話題が多く, 非被災地域であるほど災害支援に関する話題が多いといった特徴があるとしている。

これらのように, 災害時に被災地域から発信された情報が多く活用されていることがわかる。

山口らは, ソーシャルストリームから検出されたローカルイベントを用いて, ユーザの位置推定を行う方法を提案している[7]。Twitterにおけるソーシャルストリームを用いて評価実験を行い, 検出されたイベントの妥当性を示し, ユーザの位置推定の制度および効率性の点で有効性を示している。

Robら[3]のSocial Triangulationは, 地域住民は自分の住んでいる地域の公共サービスなどの地域組織のアカウントをフォローしていると仮定し, ユーザの位置を推定する方法である。Robらは, 調査した特定の市の地域組織をフォローしているユーザの中で, プロフィールに場所情報を登録しているユーザ79,998人のうち, 68%を地域住民として特定できたとしている。また, ユーザがフォローしている地域組織を分類している。1つもしくは2つの地域組織をフォローしているユーザはTVなどの報道メディアをフォローしている場合が多く, プロフィールに場所を登録している可能性が低いとしている。10以上の地域組織をフォローしているユーザは, プロフィールに場所を登録している可能性が高く, その場所が市単位で特定できるとしている。

日本では, Twitterを使って被災地の情報が発信されるこ

[†]和歌山大学システム工学部, Faculty of Systems Engineering, Wakayama University

^{*1} <https://twitter.com>

^{*2}自治体における災害時の情報発信と収集に向けて(Twitter Japan発表資料)

http://www.kantei.go.jp/jp/singi/it2/senmon_bunka/bousai_SNS_kentoukai/dai2/shiryo_3_5.pdf

^{*3}熊本地震: ツイッター投稿, 1週間で2610万件 - 毎日新聞:

<http://mainichi.jp/articles/20160519/k00/00m/040/059000c>

^{*4}Twitter ライフライン (@TwitterLifeline), <https://twitter.com/TwitterLifeline>

とが多いため、Rob らの方法を用いることで、より多くのユーザの場所を推定することができ、災害時により多く活用できる情報を抽出できるようになると考えられる。また、Twitter ライフライン (@TwitterLifeline) で各都道府県ごとのリストがまとめられており、これらのリストを用いてデータセットを作成することで、Rob らの分析よりも広範囲での分析をできると考えられる。

3. データセット

3.1 公共アカウント

Twitter ライフライン (@TwitterLifeline) のリストでまとめられている各都道府県の災害時に役に立つアカウント一覧から、アカウントを取得した。このリストに含まれるアカウントには、各市区町村、各市区町村の防災課、地域メディアなどが含まれている。Twitter ライフラインのリストには 568 名のアカウント (以降、公共アカウントとする) が登録されている。アカウントは、Twitter REST API の lists/list⁴¹ および lists/members⁴² を用いて取得した。

3.2 公共アカウントのフォロワー

3.1 節の公共アカウントのフォロワーを、Twitter REST API の followers/ids⁴³ および followers/list⁴⁴ を用いて取得した。取得した全てのアカウントのフォロワー数は 10,593,450 名、そのうちユニークなフォロワー数は 5,791,644 名であった。

表 1 に、各都道府県の公共アカウント数、各都道府県の公共アカウントの合計フォロワー数、およびユニークなフォロワー数を示す。

また、ユニークなフォロワー 5,791,644 名のうち、プロフィールの場所に情報を登録しているユーザは 1,673,570 名であった。このプロフィールの場所の情報は、「日本 東京」「和歌山市」のような地名だけでなく、「世界の果て」や「みかん県」など地名と関係のない文字列が含まれている場合、特定の地名と識別できない場合がある。

4. フォロワーの場所推定

4.1 推定方法

Rob らは、ユーザがプロフィールに登録している場所情報の曖昧性を少なくするために、Google マップおよび Google Fusion Tables を用いている [3]。しかし、Rob らが検証した 79,978 名に対して、プロフィールに場所を登録しているユニークフォロワー数が多いため、本稿では日本語形態素解析システム JUMAN++⁴⁵ および JUMAN++ の Python の

表 1: 各都道府県の公共アカウント

都道府県	アカウント数	フォロワー数	
		合計	ユニーク
北海道	29	247,599	172,800
青森県	13	157,210	102,471
岩手県	14	206,153	120,950
宮城県	14	441,926	329,735
秋田県	10	92,001	74,544
山形県	9	93,216	69,292
福島県	11	469,767	354,799
茨城県	28	331,365	199,559
栃木県	12	167,457	149,688
群馬県	10	42,480	34,595
埼玉県	39	341,030	232,367
千葉県	20	283,153	233,689
東京都	58	4,282,901	2,852,155
神奈川県	25	668,515	397,608
新潟県	13	209,140	130,887
富山県	7	20,423	15,537
石川県	4	15,885	14,168
福井県	5	6,843	6,829
山梨県	10	85,533	71,527
長野県	9	80,740	68,483
岐阜県	5	5,690	5,640
静岡県	9	187,304	130,435
愛知県	10	171,056	152,904
三重県	6	34,346	29,597
滋賀県	3	21,112	21,347
京都府	7	229,672	160,447
大阪府	40	241,882	182,565
兵庫県	13	198,964	174,191
奈良県	7	37,655	33,724
和歌山県	11	32,551	24,385
鳥取県	7	17,834	12,252
島根県	5	17,636	13,726
岡山県	7	21,214	18,247
広島県	4	104,841	97,880
山口県	5	10,716	9,877
徳島県	5	29,851	27,639
香川県	5	21,777	17,192
愛媛県	4	4,717	4,267
高知県	3	21,916	21,619
福岡県	11	113,739	106,346
佐賀県	8	63,123	54,596
長崎県	10	46,650	36,617
熊本県	10	221,574	141,892
大分県	9	89,217	71,757
宮崎県	2	22,767	22,877
鹿児島県	16	95,835	71,996
沖縄県	5	110,080	107,041

ラッパーである PyKNP⁴⁶ を用いて、プロフィールに登録している場所情報に対して形態素解析を行った。

形態素解析の結果として、品詞細分類が「地名」となった形態素を、「都道府県」「郡」「市区町村」「不明」の 5 つの項目に分類する。分類方法は以下の通りで、地名が確定できた時点で分類を終了する。

- 1) 次の形態素に「都/府/県」「郡」「市/区/町/村」のいずれかの地名の接尾辞がある場合は、「地名 + 地名の接尾

⁴¹GET lists/list, Twitter Developer, <https://developer.twitter.com/en/docs/accounts-and-users/create-manage-lists/api-reference/get-lists-list>

⁴²GET lists/members, Twitter Developer, <https://developer.twitter.com/en/docs/accounts-and-users/create-manage-lists/api-reference/get-lists-members>

⁴³GET followers/ids, Twitter Developers, <https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-followers-ids>

⁴⁴GET followers/list, Twitter Developers, <https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-followers-list>

⁴⁵JUMAN++, KUROHASHI-KAWAHARA LAB, <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

⁴⁶PyKNP, KUROHASHI-KAWAHARA LAB, <http://nlp.ist.i.kyoto-u.ac.jp/index.php?PyKNP>

辞」として地名を確定し、確定した地名の項目に分類

2) 意味情報が「自動獲得:Wikipedia Wikipedia 複合地名 Wikipedia 多義」の場合は、同義語の数によって、以下のように分類して 3) の処理に進む

- 同義語が 1 つの場合は、その意味情報を取得
- 同義語が 2 つでそれぞれの意味情報に県名と市名が一緒の場合は、その市名の意味情報を取得。それぞれの意味情報が異なる場合は、同義語の意味情報をすべて取得
- 同義語が 3 つ以上の場合は、同義語の意味情報をすべて取得

3) 意味情報に「都/府/県」「郡」「市/区/町/村」のいずれかの地名の接尾辞がある場合は、「地名+意味情報から取り出した地名の接尾辞」として地名を確定、確定した地名の項目に分類

4) どこにも分類できない場合は、「不明」に分類

ただし、「和歌山県」は「和歌山」「県」の形態素に分割されるが、「北海道」は「北海」「道」と形態素が分割されず、「北海道」となるため、「北海道」という形態素の時点で地名を確定している。

場所の情報を形態素解析した場合に、品詞細分類が「地名」となる形容詞が複数含まれることがある。その場合は、以下のように分類している。

- 形態素が「郡」「市区町村」に分類できる場合は、意味情報に都道府県名も含まれるため、都道府県名も登録する。
- 地名を各分類に登録するとき、すでに登録されている地名と都道府県名、郡名、市区町村名が一致しない場合は、すでに登録されている地名も含め地名候補とし、地名は確定できないとする。

また、プロフィールの場所情報は、文字の種類によって「和歌山」「わかやま」「ワカヤマ」「Wakayama」のように表記揺れがある。形態素解析をする前に、場所の情報は以下のように加工している。

- 一部のカタカナが含まれる地名をのぞいて、カタカナのまま形態素解析を行うと、JUMAN++ で品詞細分類が「地名」ではなく、「カタカナ」になってしまうため、半角カタカナ、およびカタカナはひらがなに変換
- 全角英字、およびローマ字の長音は小文字の半角英字に変換
- ローマ字は JUMAN++ で品詞細分類が「地名」ではなく、「アルファベット」になってしまうため、ローマ字を漢字に変換

ローマ字表記から漢字表記への変換は、全国地名のローマ字表記データ¹⁾をもとにローマ字表記と漢字表記の地名の対応表を作って変換をしている。

¹⁾yanok.net, <http://yanok.net/dist/romaji-chimei-csv/>

表 2: ユニークフォロワーの場所推定 (都道府県別, 上位 5 件)

都道府県	推定数	人口 (千人)	推定数/人口 (%)
東京都	246,042	13,515	1.82
神奈川県	71,805	9,126	0.79
埼玉県	51,455	7,267	0.71
大阪府	49,429	8,839	0.56
千葉県	42,976	6,223	0.69

表 3: ユニークフォロワーの場所推定 (都道府県別, 下位 5 件)

都道府県	推定数	人口 (千人)	推定数/人口 (%)
島根県	3,195	694	0.46
石川県	3,066	1,154	0.27
鳥取県	2,708	573	0.47
宮崎県	2,341	1,104	0.21
福井県	2,138	787	0.27

4.2 ユニークフォロワーの場所推定

3.2 節のプロフィールに場所情報を登録しているユニークフォロワー 1,673,570 名の場所の推定を行った。

表 2, および表 3 に、都道府県別の推定結果を示す。表 2, および表 3 の「推定数」は、全ユニークフォロワー 1,673,570 名の中で、登録されている都道府県のリストに関係なく、推定された都道府県の数を示す。表 2, および表 3 の「人口」は、総務省統計局の「都道府県別人口と人口増減率」²⁾のデータを使用している。表 2, および表 3 の「推定数/人口」は、推定数と人口の割合を表している。表 2 より、人口が多いほど推定数のが多く、表 3 より、人口が少ないほど推定数が少ないことがわかる。これより、公共アカウントのフォロワーは、各都道府県の人口に応じて、プロフィールに推定可能な場所情報を登録している割合が変動することがわかった。

4.3 都道府県との一致数

表 4 に、都道府県の一致数を示す。表 4 の「ユニークフォロワー数」は、表 1 の「ユニークフォロワー数」と同じである。表 4 の「一致数」は、都道府県のリストと場所を推定したユーザの都道府県が一致したものである。4.2 節の人口比と異なり、一致数/ユニークフォロワー数の比は人口に関係がなかった。

全体の一致数/ユニークフォロワー数は約 8.8% だった。Rob らは、調査した特定の市の地域組織をフォローしているユーザの中で、プロフィールに場所情報を登録しているユーザ 79,998 人のうち、68% を地域住民として特定できたとしている [3]。この Rob らの 68% に対して、今回の結果は大きく下回った。この原因としては、全体のユニークフォロワー 5,791,644 名のうち、プロフィールに場所情報を登録してい

²⁾日本の統計 2018 - 第 2 章 人口・世帯, 統計局ホームページ, <http://www.stat.go.jp/data/nihon/02.html>

表 4: 都道府県の一致数

都道府県	ユニーク フォロワー数	一致数	一致数/ユニーク フォロワー数 (%)
北海道	172,800	31,581	18.3
青森県	102,471	9,727	9.5
岩手県	120,950	12,438	10.3
宮城県	331,948	17,253	5.2
秋田県	76,520	7,077	9.2
山形県	70,586	3,274	4.6
福島県	354,837	9,946	2.8
茨城県	199,559	25,962	13.0
栃木県	149,688	9,572	6.4
群馬県	34,595	6,052	17.5
埼玉県	232,367	30,722	13.2
千葉県	233,689	29,448	12.6
東京都	2,852,155	197,288	6.9
神奈川県	397,608	51,061	12.8
新潟県	130,887	15,220	11.6
富山県	15,537	2,557	16.5
石川県	14,168	1,824	12.9
福井県	6,829	998	14.6
山梨県	71,527	5,784	8.1
長野県	68,483	9,862	14.4
岐阜県	5,640	963	17.1
静岡県	130,703	15,793	12.1
愛知県	156,869	20,753	13.2
三重県	30,357	3,063	10.1
滋賀県	23,658	2,989	12.6
京都府	166,142	15,214	9.2
大阪府	220,771	26,864	12.2
兵庫県	182,231	13,842	7.6
奈良県	35,828	2,441	6.8
和歌山県	26,065	2,462	9.4
鳥取県	13,015	1,780	13.7
島根県	14,431	2,118	14.7
岡山県	33,379	2,960	8.9
広島県	105,735	14,459	13.7
山口県	10,698	1,707	16.0
徳島県	28,221	2,051	7.3
香川県	18,135	2,324	12.8
愛媛県	5,361	966	18.0
高知県	22,406	3,093	13.8
福岡県	111,636	11,606	10.4
佐賀県	56,284	2,633	4.7
長崎県	38,462	3,233	8.4
熊本県	150,659	10,404	6.9
大分県	73,743	4,180	5.7
宮崎県	23,951	1,211	5.1
鹿児島県	74,270	5,893	7.9
沖縄県	109,618	6,277	5.7
合計	7,505,472	658,925	-
平均	159,691	14,020	8.8

るフォロワーが約 30%の 1,673,570 名と少なく、都道府県を推定することができるフォロワーが約 16.1%の 933,841 名とさらに少なかったためだと考えられる。

表 5 に、都道府県の一致数を示す。表 5 の「場所情報登録フォロワー数」は、表 1 の「ユニークフォロワー数」のうち、プロフィールに場所情報を登録しているフォロワーである。表 5 の「一致数」は、都道府県のリストと場所を推定したユーザの都道府県が一致したものである。

全体の一一致数/場所情報登録フォロワー数は、約 28.9%であった。表 5 において、最もこの割合が高かったのは愛知県で約 44.4%、最も低いのは福島県で約 8.7%であった。福島県に関しては、JUMAN++ で「福島」を解析した結果、複数の地名候補があるために、地名が推定できたものが少なかったためと考えられる。

プロフィールに場所情報を登録しているユーザの中であれば、より多くのユーザをその都道府県の地域のユーザである推定できることがわかった。プロフィールの場所情報に登録はしており、JUMAN++ で地名と判定されたが、今回の推定方法で地名と推定できなかったものが 227,190 件あり、全体 (2,281,307 件) の約 10%あった。プロフィールの場所情報は登録されない場合も多く、JUMAN++ で地名と判断しても特定の地名と推定できない場合があるため、ツイートのジオタグなども併用する必要があると考えられる。

5. おわりに

本稿では、公共アカウントをフォローしていることによって、ユーザの位置推定ができるかを検証した。Twitter ライフライン (@TwitterLifeline) のリストでまとめられている各都道府県の災害時に役に立つアカウント一覧から、各都道府県の公共アカウントを取得し、その公共アカウントのフォロワーがその地域のユーザであるかを確認した。

各都道府県において、地域住民と推定できたのはユニークユーザ中だと約 8.8%、ユニークユーザの中でもプロフィールに場所情報を登録しているユーザ中だと約 28.9%と、Robらが 68%を地域住民として特定できたしている [3] のに対して下回っていた。ユーザのプロフィールの場所情報をもとに推定したが、全体のユニークなユーザ数に対して、プロフィールに場所を登録しているフォロワー数が低かったためだと考えられる。今後の予定として、他の方法で用いられているツイートのジオタグを用いる方法などを用いて、地名の推定可能な量を増やし、地域組織のフォロワーを地域住民として推定できるか確認を行う。

参考文献

- [1] 山本太郎, 橋元良明, 中村功, 関谷直也, 小笠原盛浩, 千葉直子, 関良明, 高橋克巳: Twitter 利用を中心とする震災時の情報行動と通信不安 - 関東 Twitter 利用者ウェブ調査, 東京大学大学院情報学環情報学研究 調査研究編, Vol.28, pp.115-160 (2012) .
- [2] 榊剛史, 原久美子, 吉田光男, 鳥海不二夫, 篠田孝祐, 栗原聡, 風間一洋, 野田五十樹: 災害情報基盤構築に向けたテキストデータからの地理情報抽出システム, 第 28 回人工知能学会全国大会, 1H2-NFC-02a-4, pp.1-4 (2014) .
- [3] Rob Grace, Jess Kropczynski, Scott Pezanowski, Shane Halse, Prasanna Umar, Andrea Tapia: Social Triangulation: A new method to identify local citizens using social media and their local information curation behaviors, Proceedings of the 14th ISCRAM Conference, pp.902-915(2017).
- [4] 宮部真衣, 荒牧英治, 三浦麻子: 東日本大震災における Twitter の利用傾向の分析, 情報処理学会研究報

表 5: 場所情報を登録しているフォロワーと一致数

都道府県	場所情報登録 フォロワー数	一致数	一致数 / 場所情報登録 フォロワー数 (%)
北海道	73,966	31,581	42.7
青森県	38,927	9,727	25.0
岩手県	53,691	12,438	23.2
宮城県	106,850	17,253	16.1
秋田県	26,731	7,077	26.5
山形県	22,604	3,274	14.5
福島県	114,905	9,946	8.7
茨城県	74,217	25,962	35.0
栃木県	39,112	9,572	24.5
群馬県	15,076	6,052	40.1
埼玉県	78,496	30,722	39.1
千葉県	74,213	29,448	39.7
東京都	646,951	197,288	30.5
神奈川県	132,348	51,061	38.6
新潟県	50,588	15,220	30.1
富山県	7,423	2,557	34.4
石川県	6,884	1,824	26.5
福井県	3,128	998	31.9
山梨県	23,097	5,784	25.0
長野県	25,991	9,862	37.9
岐阜県	2,328	963	41.4
静岡県	42,059	15,793	37.5
愛知県	46,737	20,753	44.4
三重県	17,264	3,063	17.7
滋賀県	10,458	2,989	28.6
京都府	52,407	15,214	29.0
大阪府	79,009	26,864	34.0
兵庫県	59,480	13,842	23.3
奈良県	18,743	2,441	13.0
和歌山県	12,543	2,462	19.6
鳥取県	6,924	1,780	25.7
島根県	6,368	2,118	33.3
岡山県	11,654	2,960	25.4
広島県	32,996	14,459	43.8
山口県	5,199	1,707	32.8
徳島県	14,099	2,051	14.5
香川県	8,963	2,324	25.9
愛媛県	2,404	966	40.2
高知県	8,651	3,093	35.8
福岡県	35,552	11,606	32.6
佐賀県	19,882	2,633	13.2
長崎県	17,684	3,233	18.3
熊本県	53,880	10,404	19.3
大分県	24,736	4,180	16.9
宮崎県	10,263	1,211	11.8
鹿児島県	25,324	5,893	23.3
沖縄県	40,502	6,277	15.5
合計	2,281,307	658,925	-
平均	48,538	14,020	28.9

ルメディア間話題の比較分析, 情報処理学会研究報告, Vol.2013-ICS-170, No.7, pp.1-5 (2013).

- [7] 山口祐人, 伊川洋平, 天笠俊之, 北川博之: ソーシャルメディアにおけるローカルイベントを用いたユーザー位置推定手法, 情報処理学会論文誌 データベース, Vol.6, No.5, pp.23-37 (2013).

告, グループウェアとネットワークサービス研究会, Vol.2011-GN-81, No.17, pp.1-7 (2011).

- [5] Son Doan, Bao-Khanh Ho Vo, and Nigel Collier: An analysis of Twitter messages in the 2011 Tohoku Earthquake, In 4th ICST International Conference on eHealth, pp.58-66 (2011).
- [6] 小川祐樹, 野田五十樹, 山本仁志, 後藤真太郎, 和崎宏, 五味壮平, 鳥海不二夫: 災害時におけるソーシャ