

THE IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS (JAPANESE EDITION)

IEICE | **電子情報通信学会**
D | **論文誌** 情報・システム

VOL. J102-D NO. 9
SEPTEMBER 2019

本PDFの扱いは、電子情報通信学会著作権規定に従うこと。
なお、本PDFは研究教育目的（非営利）に限り、著者が第三者に直接配布することができる。著者以外からの配布は禁じられている。

情報・システムソサイエティ

一般社団法人 **電子情報通信学会**

THE INFORMATION AND SYSTEMS SOCIETY

THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

SNS を用いた言語表現パターン分析による降雨強度レベル推定手法

佐野 仁美^{†,††a)} 伊藤 正彦^{††,†††*} 生駒 栄司^{††} 喜連川 優^{††,††††}
小口 正人[†]

Rainfall Intensity Level Estimation Method by Analysis of Word Expression
Pattern Using SNS

Hitomi SANO^{†,††a)}, Masahiko ITOH^{††,†††*}, Eiji IKOMA^{††},
Masaru KITSUREGAWA^{††,††††}, and Masato OGUCHI[†]

あらまし 近年、国内では局地的降雨の事象が多く発生し、気象への関心が高まっている。こうした気象状況は物理的観測情報のほか、投稿情報の集合知から得る試みもある。投稿情報による表示は、現状、「晴れ・くもり・雨」等の天気区分で表すのが主流だが、降雨の程度（降雨強度）が把握可能となれば、防災上の観点から有用な情報を得ることができる。そこで本研究では、言語表現から降雨強度の推定を行う手法を提案する。本研究では、代表的な SNS の一つである Twitter の発信情報と国内で最も高解像度な観測情報である XRAIN との数値的相関を解析し、言語表現と降雨強度の相関及び降雨強度の季節的変動の特徴を検出し、言語表現パターンと季節（月）を組み合わせた推定モデルを構築した。そして、本モデルを実際の Twitter 情報に適用した結果、推定値は XRAIN の実績と年間平均 87%以上で一致し、本提案手法が言語表現から物理情報相当値を推定する手法の一つとして有用であることを示した。本論文では本提案手法と解析結果を紹介する。

キーワード Twitter, SNS, XRAIN, 防災, 降雨情報

1. ま え が き

近年、国内では局地的降雨の事象が多く発生し、降雨情報の迅速な把握と共有は重要な課題となっている。局地的降雨の検知に向く気象観測情報では、国土交通省が全国に設置した X-band MP レーダによる観測情報（最小観測面積 250m メッシュ、配信は 1 分ごと）が国内で最も高解像度かつ高頻度な観測情報

を提供する。このため我々は、国内最大規模の地球環境情報プラットフォームであるデータ統合・解析システム（DIAS^(注1)）上に X-band MP レーダによる観測網（以下、「XRAIN^(注2)」と呼ぶ。）の情報共有を目的としたデータ提供基盤を開発し、2014 年 7 月から運用を開始した [1]~[3]。一方、最近ではこうした物理情報によらず、人の発信情報により天気を表示するサービスが広く活用されるようになってきている。現状は「晴れ・曇り・雨」等の表示が一般的で、降雨の強さを把握可能な研究及びシステムは我々の知る限り存在しない。しかし、言語表現から降雨の 1 時間当たりの雨量（以下、「降雨強度」^(注3)と呼ぶ。）を把握できれば、防災上の観点から有用性が高い情報を物理情報以外から得ることができる。したがって、本研究では言語表現から降雨強度を推定する手法を提案し有用性を検証する。

[†] お茶の水女子大学, 東京都
Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo,
105-0123 Japan

^{††} 東京大学, 東京都
The University of Tokyo, 4-6-1 Komaba, Meguro-ku,
Tokyo, 153-8505 Japan

^{†††} 情報通信研究機構, 小金井市
National Institute of Information and Communications
Technology, 4-2-1 Nukui-Kitamachi, Koganei-shi, 184-
8795 Japan

^{††††} 国立情報学研究所, 東京都
National Institute of Informatics, 2-1-2 Hitotsubashi,
Chiyoda-ku, Tokyo, 101-8430 Japan

* 現在, 北海道情報大学

a) E-mail: sano@tkl.iis.u-tokyo.ac.jp
DOI:10.14923/transinfj.2019JDP7010

(注1) : Data Integration and Analysis System,

<http://www.diasjp.net/>

(注2) : 国土交通省 川の防災情報 レーダ雨量「XRAIN とは」,
<https://www.river.go.jp/kawabou/reference/index09.html>

(注3) : 気象庁「気象観測の手引き」, https://www.jma.go.jp/jma/kishou/known/kansoku_guide/tebiki.pdf

表 1 降雨強度レベル
Table 1 Level of rain intensity.

強度レベル	合成雨量の範囲
1	弱い雨： 10 mm/h 未満 (0 を除く)
2	やや強い雨： 10 mm/h 以上 20 mm/h 未満
3	強い雨： 20 mm/h 以上 30 mm/h 未満
4	激しい雨： 30 mm/h 以上 50 mm/h 未満
5	非常に激しい雨，猛烈な雨： 50 mm/h 以上

本研究では言語表現の解析に代表的な SNS の一つである Twitter を利用した。ソーシャルメディアに関する研究機関^(注4)の最新動向調査^(注5)によると、現時点の国内の月間アクティブユーザ数では Twitter が約 4,500 万人と最も多く、続いて Facebook が 2,800 万人、Instagram が約 2,000 万人である。人による発信情報にはゆらぎ（曖昧性）が存在するため、解析には統計的観点から多くのデータを必要とする。したがって本研究では、より多くの情報を取得する観点から Twitter を選択した。

本研究では、Twitter を利用した発信情報（以下、「Tweet」と呼ぶ。）に含まれる言語表現から降雨情報を推定する手法を提案する。本提案手法は大別して以下の 2 部により構成される。

- ① 対象 Tweet の特定
- ② 言語表現から物理情報を推定するモデルの提案

まず、①については、先行研究を利用して Tweet に推定位置情報を付加し、位置情報付 Tweet の対象を拡大する。また、独自に構築したアルゴリズムにより発信時刻と内容を精査する。

次に、②の概要を述べる。本研究の対象は関東地域とし、推定値は気象庁による雨の強さと降り方の関係^(注6)を基準として表 1 に示す 5 段階のレベルで表す。

まず、「小雨」「豪雨」等、降雨強度を推定可能な言語表現（以下、「推定言語表現」と呼ぶ）を特定し、これらを含む Tweet に対応する XRAIN と照合した。この結果、「小雨」「霧雨」のように時間や季節によらず一定の降雨強度を示す表現や、「豪雨」「大雨」のように時間または季節により降雨強度が変動する表現など、推定言語表現と降雨強度に相関を検出できた。こ

のため、推定言語表現を降雨強度の変動幅に応じ三つの Category に分類し、当該 Category ごとに推定を行うこととした。

降雨には弱雨から強雨に推移する時間的な変動と季節変動がある。このため、我々は双方の変動の特徴を過去の Tweet 及び XRAIN の実績を用いて解析し、Category ごとに XRAIN から最適解を得る方法を定め、当該 Category 内の「推定言語表現」と「月」を要素として降雨強度レベルを推定するモデルを構築した。

そして、本提案手法を 2017 年の事例に適用した結果、推定値と XRAIN の実績は年間平均 87%以上の精度で一致した。したがって、本提案手法は言語表現から物理情報相当の値を得る有用な手法の一つと考えられる。

本論文の構成は以下のとおりである。まず 2. では関連研究及び類似サービスを紹介し、3. では本研究の対象とする Tweet の特定手法を示す。次に、4. では Tweet から降雨強度を推定する提案モデルの詳細及び解析結果を紹介する。そして 5. で適用結果と考察を述べ、6. でまとめを述べる。

2. 関連研究

SNS 情報と実データを関連付ける研究 [4] では、Calton Pu らによる物理情報（地震と降雨）と SNS 情報（Twitter, Instagram, YouTube）を統合したシステムの開発 [5]、地震発生を示す Tweet から高い確率で地震発生を検出する研究 [6]、鉄道の遅延等に関する Tweet から当該事象を検知する研究 [7]、自然災害の発生を検知しネットワーク切替を試みる研究 [8]、及び Tweet 内の言語表現から体感温度を取得する研究 [9] 等がある。

降雨と Tweet の関連研究 [10]~[20] では、「浸水・冠水」の関連 Tweet から災害検知を行う研究 [10]~[12]、降雨の関連 Tweet から当該降雨の持続時間検出を行う研究 [13]、Twitter から天気（晴れ、曇りまたは雨）を把握する研究 [14]、局地的降雨のバースト現象を捉える研究 [15], [16]、土砂災害の前兆を捉える研究 [17]、気象庁の降雨データと「雨」「地滑り」の単語の関連性を述べた研究 [18]、災害支援が必要な地点の検出を行う研究 [19] 等がある。これらの研究は、Tweet の集合知から特定事象の把握や検出を試みている。一方、個別 Tweet を議論する研究では、ジオタグ付きの Tweet から当該時点の状況を抽出する研究 [20] がある。当該研究では、Tweet から降雨有無を把握可能であり、位

(注4)：ソーシャルメディアラボ、<https://gaiax-socialmedialab.jp/>

(注5)：「2018 年 6 月更新！ 11 のソーシャルメディア最新動向データまとめ」、<https://gaiax-socialmedialab.jp/post-30833/>

(注6)：気象庁「雨の強さと降り方」、https://www.jma.go.jp/jma/kishou/known/yougo_hp/amehyo.html

置情報を正確にする必要性を述べている。本研究ではこれを参考にし、ジオタグ付き Tweet に加えて先行研究により位置情報を付加した Tweet を対象とした。

投稿情報から天気を表示する試みでは、ヤフー株式会社が運営するポータルサイト Yahoo! JAPAN 上の「みんなで実況・今の天気」^(注7)、及び株式会社ウェザーニューズが運営する「ウェザリポート」^(注8)がある。「みんなで実況・今の天気」は、登録ユーザが投稿した天気区分（晴れ、くもり、雨、雪）を直近1時間で集計し10分ごとに公開する。「ウェザリポート」では、ユーザによる投稿画像や天気区分をリアルタイムで共有する。両サービスとも投稿及び集計は市町村区の比較的広域な単位であり、天気情報は設定区分（晴れ・くもり・雨等）に従うため降雨強度までは表示されない。

3. 対象 Tweet の特定

本研究では、降雨情報を正確に得るため、地点と時点が特定可能な Tweet を対象とする。しかし、位置情報を把握可能なジオタグ付き Tweet は非常に数が少ない。また、Tweet の示す内容も「小降りだった（過去）」、「猛烈な雨が降っている（現時点）」、「雨が降りそう（未来）」など、当該 Tweet の発信時刻における事象を示すとは限らない。したがって我々は、先行研究を利用して位置情報付き Tweet の対象数を拡張するとともに、独自のアルゴリズムを構築し、Tweet の示す降雨情報の時点を特定する。本章では当該特定方法について述べる。

3.1 発信地点の設定

発信地点に正確性を求める場合、一般的にはジオタグ付き Tweet の利用が適する。しかし、これらは Tweet 全体の1%未満であり、非常に数が少ない[21]。このため、本研究では、伊藤・吉永らによる先行研究[22],[23]を利用してジオタグがない Tweet に「推定位置情報（緯度経度）」を付加し、対象 Tweet を拡大した。

伊藤・吉永らの手法はマイクロブログ（Tweet）ストリーム中の投稿中に現れる単語を元に、Tweet に推定位置情報を付加する。本手法は、まず、Tweet がジオタグ付か否かで位置情報の付加方法を大別し、ジオタグ付の場合は当該位置情報を推定位置情報に設

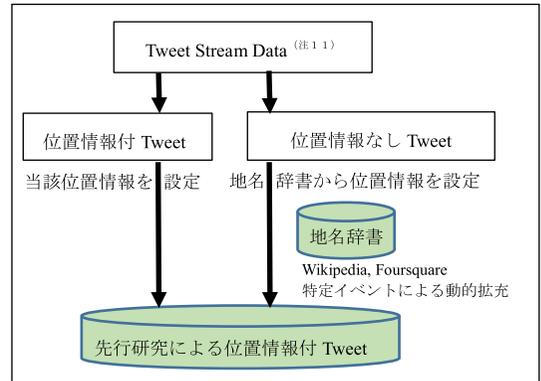


図1 推定位置情報付加の概要

Fig. 1 Outline of adding estimated position information.

定する。一方、ジオタグがない場合は、地名辞書に基づき位置情報を推定する。本手法で利用する地名辞書は、Wikipedia^(注9)に含まれる地名と位置情報に加え、Foursquare^(注10)のような位置情報サービスを介して投稿された Tweet から得た施設名と位置を組み合わせ構築している。加えて、特定の位置と強く関連付けられる名詞も都度追加する。例えば、あるイベントが特定の場所で開催された場合、Tweet 内に当該イベント名が含まれれば、場所情報を含むか否かにかかわらず、イベント開催地点を推定位置情報に設定する。

ただし、本研究では照合対象である XRAIN が極めて高解像度であるため、Tweet の位置情報誤差を勘案し、区市町村名に基づき推定位置情報が付加された Tweet は除外した。この手法の全体概要を図1に示す。

2017年のデータを使用した適用事例では、本研究の対象 Tweet をジオタグ付き Tweet のみの利用に比べ約4倍に拡大することができた。

3.2 発信時点及び内容の特定

本研究で使用する Tweet は、「現時点の降雨を表現しノイズが排除された Tweet」（以下、「有効 Tweet」と呼ぶ。）とする。以下、有効 Tweet の特定方法を説明する。

(注9) : <https://www.wikipedia.org>

(注10) : <https://ja.foursquare.com/>

(注11) : 東京大学、喜連川・豊田研究室で2011年3月11日からTwitter APIを用いて収集したツイートデータを用いている。ツイートの収集は30人の著名Twitterユーザを初期ユーザセットとして過去の投稿を取得するところから開始している。その後、彼らのタイムライン上でリツイートやメンションが行われたユーザを新たな対象としてユーザセットに加え、同様に過去の投稿を取得する。このように、ユーザセットの拡大とタイムラインのトラッキングを繰り返し、2018年11月18日時点で約574億ツイートを蓄積している[23]。

(注7) : https://www.yahoo-help.jp/app/answers/detail/p/616/a_id/44565/faq/pc-detail. (2018年8月20日時点)

(注8) : <http://weathernews.jp/report/>. (2018年8月20日時点)

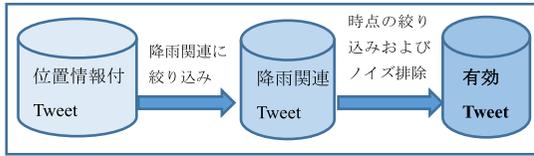


図 2 有効 Tweet の抽出イメージ

Fig. 2 Processing Image of Target-Tweet Extraction.

まず、位置情報付 Tweet から次の降雨関連用語を含む Tweet (以下、「降雨関連 Tweet」と呼ぶ。)を抽出する。

雨 (豪雨, 大雨等の雨を含む単語), 土砂降り, どしゃ降り, 夕立, スコール, 小降り, 本降り, ザーザー降り, 台風, 雹 (ひょう), 霰 (みぞれ)

次に、降雨関連 Tweet から時点や内容に関する絞り込みを行い、有効 Tweet を特定する。この処理イメージを図 2 に示す。

降雨関連 Tweet から有効 Tweet を抽出する手法は、2015 年 1~3 月の関東地域の降雨関連 Tweet を教師データとして構築した独自のアルゴリズムにより行う。具体的には、次のいずれかの条件を満たす Tweet を有効 Tweet とし、いずれも満たさない Tweet は除外する。

- ① 現在進行形の降雨を示す動詞の表現。
 - 降っている, 降ってきた, 降り出した, 降り続けている, 降り続く, 止みそう, 止んできた等 (これらの活用形も含む)。
- ② 降雨関連用語と記号 (「」以外) の組み合わせ。
 - 店を出たら雨。(「雨」+ 記号「.」) 豪雨! (「豪雨」+ 記号「!」) 等。
- ③ 降雨関連用語と助動詞や助詞の組み合わせ。
 - どしゃ降りです。等。
- ④ 降雨関連用語, 形容詞, 記号の組み合わせ。
 - 雨がすごい, すごい雨! 等。

本アルゴリズムは、過去 (「降った」「降っていた」), 未来 (「降りそう」「降るかも」等) の表現を除外する。また、推測 (「~らしい」「~みたい」等) も除外し、例えば東京で九州地区の豪雨を呟く等、Tweet の内容が発信地点を示さない事例を含まない。更に、映画, 本, 人名等の一般固有名詞を示す例も除外する。本アルゴリズムを適用した結果を表 2 に示す。データはアルゴリズム構築に用いた学習用データ (2015 年 1~3

表 2 2015 年 8 月の有効 Tweet の精度

Table 2 Accuracy of Effective Tweet (August, 2016).

区分	件数	誤件数	精度(%)
有効 Tweet	1281	42	96.7
除外 Tweet	17034	258	98.5

月)とは別の 2015 年 8 月の降雨関連 Tweet を用いた。

表 2 のとおり、有効 Tweet が誤って除外された事例は 42 件、逆に、除外 Tweet に誤って有効 Tweet が混在した事例は 258 件で、双方とも精度は 96% 超となった。この結果から、本アルゴリズムは高精度で有効 Tweet を抽出可能と考える。

4. 降雨強度推定モデルの構築

本章では、推定言語表現 (降雨強度を推定可能と考えられる言語表現) から降雨強度を取得するモデルの構築方法を説明する。

まず、推定言語表現 (全 81 語) を特定し、当該言語が示す降雨強度を過去の Tweet と XRAIN を照合した。この結果、推定言語表現ごとに降雨強度の変動幅が異なる特徴が検出できたため、推定言語表現を当該変動幅に応じて三つの Category に分類した。

降雨強度の変動は大別して 2 種あり、弱雨からピークを迎え収束する「時間的な変動」と、冬に小さく夏に大きい傾向がある「季節的な変動」である。したがって、推定言語表現から降雨強度のレベル値を推定する場合、両変動を考慮した仕組が必要となる。本研究では、降雨の時間的な変動を考慮し、Tweet の示す降雨として当該時間帯のいずれの値 (最大値, 平均値または頻出値) の選択が最適かを過去の Tweet 及び XRAIN を基に検証し、Category 別に当該取得方法を決定した。併せて、降雨強度レベル値の推定モデルでは、季節的な変動を考慮して設定値を「月」単位に細分化し、「推定言語表現」と「月」を要素として構築した。

4.1 推定言語表現の特定

本研究で使用使用する推定言語表現は、分類器を使用して 2015~2016 年の有効 Tweet から抽出した。

まず分類器の仕様を説明する。分類器は有効 Tweet を入力データとし、当該 Tweet を 1 件ずつ形態素解析した上で、あらかじめ登録された推定言語表現 (以下、「登録言語表現」と呼ぶ。)と照合する。当該処理概要を以下に示す。

＜処理概要＞

形態素解析された有効 Tweet の各語を配列 array[] に割り当て、下記①～④の比較を行う。以下①～④のいずれかの条件を満たす場合は当該有効 Tweet をファイル (File1 とする) に出力する。また、①～④のいずれも満たさない場合は当該有効 Tweet をファイル (File2 とする) に出力し、当該 Tweet 中の全ての名詞、形容詞、副詞を登録候補としてファイル (File3 とする) に出力する。

① array[x] の品詞 = 名詞の場合、当該名詞を登録言語表現とする。

例：豪雨、大雨、本降り、スコール

② array[x] の品詞 = 形容詞かつ前後に“雨”が存在する場合、当該形容詞を登録言語表現とする。

例：強い、弱い、凄い、酷い

③ array[x] の品詞 = 副詞かつ前後に指定の動詞 (*) が存在する場合、当該副詞を登録言語表現とする。
(*) 指定の動詞：降る、降り出す、降り続く

例：しとしと、ポツポツ、チラチラ、ザーザー

④ array[x]～array[x+3] の連結に特殊表現を含む場合、当該連結言語表現を登録言語表現とする。

例：滝のような、バケツをひっくり返す

次に、推定言語の特定手法を説明する。まず、3.2 で掲げた降雨関連用語を分類器の登録言語表現に初期登録する。それから 2015 年の有効 Tweet を分類器にかけ、上記 File3 に出力された登録言語表現の候補について、File2 へ出力された該当の Tweet における使われ方を確認しながら検証する。検証の結果、追加が必要と考えられる推定言語表現を全て分類器の登録言語表現に手動追加する。これを繰り返し、追加分がこれ以上ないと判断できるまで継続実行する。このように 2015 年の登録言語表現を確定させた後、2016 年の有効 Tweet について同様の処理を行う。ここで、2016 年の有効 Tweet により追加登録となった登録言語表現は 1 件のみであった。このため、その 2 年分の教師データを元とした登録言語表現については必要十分と判断し、推定言語表現 (全 81 の言語表現) を確定させた。

また、今回の推定言語表現は関東地域の有効 Tweet により構築したため、他地域には用いない。つまり、地域固有の表現 (方言) を勘案して、推定言語表現は地域単位での設定とする。

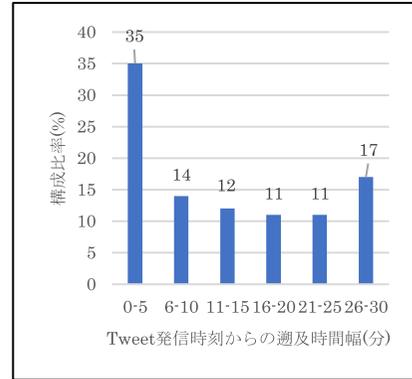


図 3 Tweet 発信時刻と降雨ピークの時間差
Fig. 3 Time Lag between Tweet Time and Rain-Peak.

4.2 Tweet 発信における入力時間差の考慮

一般的に、Tweet 発信時には入力等による時間差が生じる。このため、Tweet と照合する XRAIN は発信時刻時点の XRAIN だけでなく、この時間差を考慮した範囲のデータも検証する必要がある。我々は当該時間差の傾向把握のため、Tweet の発信時刻と降雨群のピーク時刻との差を検証した。検証方法は、2015 年の関東地域の有効 Tweet ごとに XRAIN (発信時刻から 30 分遡及) を抽出してピーク値の時刻を取得し、Tweet 発信時刻との差を 5 分単位で集計した。この集計結果が全体に占める割合 (%) を図 3 に示す。

図 3 によると、Tweet 全体の約 3 分の 1 は降雨ピークから 5 分以内の発信である。5 分以上遡ると急激に割合が減少し、以降は同様の傾向を示した。また、30 分まで遡ると別の降雨群発生の可能性も確認できた。我々は降雨強度の大きい降雨に関する Tweet は降雨のピークに反応して発信されると仮定し、当該 Tweet の示す降雨を Tweet 発信時刻から遡及 5 分間の XRAIN を基に取得することとした。

4.3 推定モデルの候補

Tweet と照合する XRAIN を遡及時間に応じた件数 (XRAIN は毎分データのため、遡及 5 分間の場合は 5 件) 取得する。その後、それらの複数の XRAIN のデータから推定言語表現と照合する 1 データを絞り込む。この選定基準は 3 パターンを検討した。具体的には遡及時間範囲の XRAIN の①最大値、②平均値、③最頻出値のいずれかを選定する。参考として、「小雨」及び「豪雨」の事例 (遡及 5 分間) を図 4 に示す。

図 4 を参照すると、「小雨」を含む Tweet に対応する XRAIN の降雨強度レベルは 5 分間全て「レベル

Tweet時刻：10:30 小雨			Tweet時刻：11:10 豪雨		
XRAIN			XRAIN		
時刻	雨量	レベル	時刻	雨量	レベル
10:30	1mm/h	1	11:10	25mm/h	3
10:29	3mm/h	1	11:09	29mm/h	3
10:28	2mm/h	1	11:08	25mm/h	3
10:27	1mm/h	1	11:07	37mm/h	4
10:26	4mm/h	1	11:06	34mm/h	4

図4 週及時間内の XRAIN の候補レベル
Fig.4 Candidate level of XRAIN within past time.

「小雨」			「豪雨」		
日付	時刻	レベル	日付	時刻	レベル
2015/08/20	05:20	1	2015/08/01	19:34	2
2015/08/20	06:44	1	2015/08/02	15:13	2
2015/08/20	08:12	1	2015/08/02	15:42	3
2015/08/20	09:29	1	2015/08/02	17:39	3
2015/08/20	14:06	1	2015/08/02	18:03	5
2015/08/20	18:43	1	2015/08/02	15:06	4
(以下、省略)			(以下、省略)		
月間構成比率：			月間構成比率：		
レベル1： 100%			レベル1： 14.7%		
			レベル2： 29.4%		
			レベル3： 32.4%		
			レベル4： 14.7%		
			レベル5： 8.8%		

図5 降雨強度レベルの推定言語表現別・月別集計
Fig.5 Monthly aggregation of rain intensity level by each words.

1]であり、この結果、Tweetごとの降雨強度レベルは①～③のいずれも「レベル1」となる。一方、「豪雨」ではXRAINの結果がバラつき、Tweetごとの降雨強度レベルは①レベル4、②レベル3、③レベル3となる。

次に、2015年8月の「小雨」「豪雨」の事例をTweetごとに①で抽出した月別集計結果を図5に示す。

図5を参照すると、「小雨」の降雨強度レベルは月内全て「レベル1」である。一方、「豪雨」の月内最大件数のレベル値は「レベル3」だが、月間構成比率（以下、「統計的精度」と呼ぶ）は約3割と低い。

次に、月内最大件数の降雨強度レベル値の統計的精度を年間で調査した（表3参照）。

表3の結果では、「小雨」の降雨強度レベルは全て「レベル1」を示し、統計的精度も96%以上である。一方、「豪雨」は、降雨レベル値に1～3の幅があり、夏に大きい値を示すなど季節変動がある。したがって我々は、言語表現から推定値を一意に求めるより、季節の

表3 降雨強度レベル値の年間傾向（2015年）
Table 3 Annual trend of frequent values (2015).

	小雨		豪雨	
	レベル	統計的精度 (%)	レベル	統計的精度 (%)
1月	レベル1	100	レベル1	50
2月	レベル1	100	レベル1	83
3月	レベル1	100	レベル1	85
4月	レベル1	100	レベル1	84
5月	レベル1	100	レベル2	77
6月	レベル1	98	レベル2	74
7月	レベル1	97	レベル3	54
8月	レベル1	96	レベル3	32
9月	レベル1	98	レベル3	46
10月	レベル1	100	レベル2	65
11月	レベル1	100	レベル1	88
12月	レベル1	100	レベル1	67
平均		99		67

表4 推定値設定パターン
Table 4 Estimated model pattern.

XRAINの選別	月内抽出方法	
	最大件数	中央値
最大値	A	B
平均値	C	D
頻出値	E	F

要素を含めた条件で細分化の方がより高精度の結果を得られると考えた。この細分化の粒度は「月」や「週」が考えられるが、「週」は細かく繁雑になるほか各週のデータが少なく季節的な降雨変動の特徴を捉えにくい。このため我々は、降雨強度の推定単位を「月」とし、降雨強度レベルを推定するモデルを「推定言語表現」と「月」を要素として構築することとした。

ここで、月内の分散したレベル値から月ごとのレベル値を選定するには次の2種の方法が考えられる。

- (i) 同月内で最大件数のレベル値
- (ii) 同月内で中央値を示すレベル値

したがって、推定言語表現ごと・月ごとに降雨強度レベル値を決定する方法（以下、「推定値設定パターン」と呼ぶ。）は、Tweet単位でのXRAINの選定方法（3パターン）と月間の推定言語表現単位のレベル値選定方法（2パターン）を組み合わせた以下A～Fの6パターンを考える（表4参照）。

加えて、推定言語表現の統計的精度により最適パターンは異なる可能性が考えられる。このため本研究では、全推定言語表現を2015～2016年の月別統計

表 5 推定言語表現の分類
Table 5 Classification of estimated words.

Category	統計的精度 (年間平均)	対象推定言語表現数
Category 1	90%以上	39 語 (小雨, 霧雨, 大粒, 大量等)
Category 2	70%以上 90%未満	26 語 (ザーザー, 横殴り, タ立ち等)
Category 3	70%未満	16 語 (豪雨, 土砂降り, 雷雨等)

的精度の年間平均により三つの Category に分割した (表 5 参照)。

表 5 の Category 1 には年間の平均降雨強度がレベル 1 の推定言語表現が全て含まれた。しかし, 同強度がレベル 2 以上の推定言語表現は Category1~3 にバラつき, 特に Category3 ではレベル 3 以上の推定言語表現が多い。この結果から, 降雨強度が小さい推定言語表現は統計的精度が高く, 降雨強度が大きくなると人間の感覚によるバラつきが多くなると言える。したがって, 本研究のモデルでは Category ごとに推定値設定パターンを決定し, 推定言語表現別・月別の設定値 (推定値) を設定する。

4.4 推定値設定パターン及び遡及時間の決定

まず, Category ごとの推定値設定パターンを決定する。教師データは 2015~2016 年の Tweet と XRAIN とし, 6 パターン全てについて, 推定言語表現及び月ごとのレベル値を設定した。また, XRAIN の遡及時間は 5 分としたが, 図 3 において遡及 10 分以内の場合に構成割合が半数を超えるため, 遡及時間を 10 分に拡大した実験も併せて行った。つまり, 実験対象モデルは推定言語表現別 (全 81 語) × 月別 (12 ヶ月) × 推定値設定パターン別 (6 種) × 遡及時間別 (2 種) となる。そして, これらのモデルに 2016 年の有効 Tweet (9,695 件) を適用し, 得られた推定値と XRAIN を照合して一致率を算出した。まず, Category ごとの一一致率の年間平均の結果を表 6 に示す (表中の値は遡及 5 分間の XRAIN による数値を示し, 同 10 分間による結果は括弧内に示す)。また, 一致率は以下の式により得る。

$$\text{一致率} = \frac{\text{月内一致件数}}{\text{月内一致件数} + \text{月内不一致件数}}$$

表 6 一致率の年間平均 (2016)
Table 6 Annual Average of Matched Rates (2016).

パターン	推定言語表現の分類 (%)		
	Category 1	Category 2	Category 3
A	88.1 (87.7)	78.8 (74.5)	69.2 (62.7)
B	86.2 (85.1)	71.2 (69.7)	56.8 (59.7)
C	97.5 (97.8)	81.6 (82.7)	79.8 (76.7)
D	92.5 (91.6)	79.2 (79.1)	69.8 (63.2)
E	99.8 (99.7)	81.7 (80.4)	71.8 (68.7)
F	99.4 (99.1)	80.8 (80.2)	65.7 (61.7)

表 7 Category ごとの推定値設定パターン
Table 7 Model pattern for each category.

Category 1	Category 2	Category 3
パターン E	パターン E	パターン C

表 8 「豪雨」の降雨強度レベル別構成比率 (2016 年 7 月)
Table 8 Appearance rate of "heavy rain" by rain intensity level.

レベル 1	レベル 2	レベル 3	レベル 4	レベル 5
6.5	13.5	42.3	13.4	24.3

表 6 より, Category ごとに最適な (一致率が最高値を示す) 推定値設定パターンを表 7 のとおり決定する。

しかし, 表 6 の全てのパターンで Category3 の一致率が最も低い。このため, Category3 の推定言語表現の統計的精度を調査した結果, 2016 年 6~8 月では同精度が低いものから順に, 「豪雨」(42.3%), 「土砂降り」(56.2%), 「すごい」(57.3%) となった。また, 不全体における件数割合は「豪雨」(42.9%) が約半数であり, 「土砂降り」(14.3%) や「すごい」(10%) などは高くない。したがって, Category3 の精度の低さは「豪雨」の不一致に起因すると考えられる。このため, 「豪雨」が頻出した 2016 年 7 月の降雨強度レベル別構成比率を調査した (表 8)。

表 8 のとおり, レベルは 1~5 に分散し, 「レベル 3」の構成比率が最も高い。ここで, 「推定言語表現が「豪雨」かつ対象月が「7 月」のレベル値を「レベル 3」とすると, 不一致となる対象が半数以上 (57.7%) となる。「豪雨」を含む Category 3 は表 7 のとおりパターン C (遡及時間内の XRAIN の平均値かつ月間集計の最大件数) でモデルを構築するが, 強度の大きい降雨

は強度の変動を繰り返し平均値は遡及時間の長さの影響を受けやすい。

このため、XRAIN の遡及時間を 0~10 分の幅に変化させ、一致率を調査した。この結果、一致率は遡及 3 分で最も高くなり、「レベル 3」の構成比率も 51% に上昇した。このため、「豪雨」の結果を Category 3 の推定言語表現に適用し、Category3 に属する推定言語表現は Tweet 発信時刻から遡及 3 分間の XRAIN によりモデル設定及び照合を行う。一方、Category1 及び Category2 は遡及 3 分間の場合も一致率がほぼ変わらなかった（遡及 5 分の一致率より微減となった）。このため、表 6 において XRAIN が遡及 5 分間の場合の一致率が 10 分間の場合より高精度となった結果も併せ、Category1 及び Category2 の同遡及時間は 5 分間とする。

4.5 Tweet 件数の計上における人口密度の影響

本手法では、Tweet を局地的情報として捉え、当該地点の XRAIN と対応付けて、関東地域全体において推定言語ごと・月ごとに集計してレベル設定を行った。しかし、集計単位を地域全体とした場合、人口が少ない県より都市部の Tweet 数が優先され、都市部の特徴を表すレベル設定となる。この場合、当該結果は人口の少ない県の特徴と合わず、結果として全体の精度を下げる可能性がある。したがって、レベル設定の際には、地域全体での集計結果と各県の人口密度を考慮した重みづけをした集計結果の比較を行うことが必要と考える。しかし、本研究の事例では、対象を関東地域としたため、両者の比較において違いが現れず、結果として集計単位は地域全体とし、人口密度による重みづけの考慮は不要と判断した。

しかし、例えば中部地域や東北地域等、太平洋側と日本海側とで気候が異なる地域では、該当地域全域での集計と人口密度の調整を行った集計とで結果が異なり、県ごとの特徴が顕現化する可能性がある。この場合は集計単位を地域別ではなく、県別あるいは緯度経度を利用した一定の区分等に細分化したレベル設定とする等の考慮が別途必要になると考えられる。したがって、県別の特徴の有無及び集計単位の適合性の事前調査を目的として、地域全体の集計と人口密度を考慮した重みづけによる集計結果の比較を行う必要がある。

このため、今回は関東地域が対象のため結果として人口密度の調整は不要となったが、今後の他地域におけるレベル設定を鑑み、参考情報として、今回関東地

表 9 人口密度に応じた按件数

Table 9 Distribution number according to population density.

県名	人口密度(人/㎡)	J_X : 調整値
東京	6168.7	1
神奈川	3777.7	1.63
埼玉	1913.4	3.22
千葉	1206.5	5.11
茨城	478.4	12.89
群馬	310.1	19.89
栃木	308.1	20.02

域のデータに対して実施した人口密度による影響確認の手順及び結果を本節に示す。

関東地域の人口密度は表 9^(注12)に示すとおり、都県間で約 20 倍の差が生じる事例がある。

そこで、統計上の信頼性の観点から、レベルごとの月間構成比率の集計時に都県の人口密度に応じた按件数で計上した場合の影響を調査した。以下に表す式のとおり、按件数 (R_X) は、Tweet の位置情報 (緯度経度) に応じた都県 X の件数 K_X に人口密度に応じた調整値 (表 9: J_X) を乗じて算出する。

$$R_X = K_X \times J_X^* \quad \text{※表 9 参照}$$

モデルを再構築した結果、再構築前のモデル設定値からの変更率は全設定値 972 件 (81 語 × 12 ヵ月) の約 6% に留まった。したがって、年間平均一致率も人口密度を考慮しない場合とほぼ変わらない。これは関東地域に気象差及び人の感覚差が少ないことが理由の一つと考えられる。この結果から、関東地域では人口密度に伴う統計的影響が少ないため考慮しない。今後別地域でモデル構築する際に改めて同調査を行い適切な集計単位 (県別等) を検討する扱いとする。

4.6 降雨強度推定モデルの構築手法

最後に、これまでの議論の結果として、本モデルの概要を総括する。本モデルは 2015~2016 年の 2 年間の有効 Tweet (計 21,452 件) 及び XRAIN を教師データとして、次の①~④の基準で構築する。

- ① 推定言語表現を統計的精度に応じて分類し、当該分類ごとに推定値設定パターンを決定する。
- ② ①の設定パターンに基づき、推定言語表現別・月別の設定値を決定する。
- ③ 教師データの XRAIN は Tweet の発信時刻から遡

(注12)：総務省統計局「平成 27 年国勢調査 (人口等基本集計結果)」、<http://www.stat.go.jp/data/kokusei/2015/kekka/kihon1/pdf/gaiyou1.pdf>

及5分間のデータを利用する。ただし、Category3の推定言語表現のみ、週及3分間とする。

④ 人口密度に伴う統計的影響は考慮しない。

5. 適用結果及び考察

5.1 2017年の事例に対する適用結果

4.で決定したモデルを2017年の事例に適用した。手順は、まず、2017年の位置情報付きTweetから降雨関連Tweet(127,641件)を抽出し、これらからXRAINの範囲内地点における有効Tweet(9,213件)を特定する。次に、これらの有効Tweetにモデルを適用して推定値をTweetごとに算出する。そして、この推定値と当該Tweetに対応するXRAINの降雨レベル推測値と照合した。この照合結果を月別の一致率として算出し、表10に示す。

表10の「一致率月間平均」を参照すると、12カ月中9カ月で80%を超え、年間平均でも87.1%で一致する高精度の結果が得られた。特に、「Categoryごとの全体に占める割合」を参照すると、統計的精度の高いCategory1の割合が高い場合に全体の精度が上昇すると考えられる。逆に、Category3の出現割合が多い6~8月は多少一致率が減少するものの70%を超えており、全体的に精度が良い結果が得られたと言える。

5.2 モデルの構成要素有無による結果の比較

本モデルの構成要素の「月」及び「Categoryごとの推定値設定パターン」を含まない場合の手法を考察し、本モデルの構成要素の必要性を評価する。具体的には、以下の実験1及び2の一致率を算出し、本モデルによる結果と比較する。

実験1:月別とせず推定言語表現ごとに一意のレベル値を設定する方法

実験2:Category別とせず共通の推定値設定パターンでレベル値を設定する方法。

実験1の手順を述べる。まず、Categoryごとの推定値設定パターン(表7)に基づき教師データから特定した降雨強度レベルを推定言語表現ごとに一意の値を設定する。例えば、Category1の事例では、2015~2016年の各Tweetから推定言語表現を取り出し、当該Tweetに対応する週及時間5分間中のXRAINの頻出値を抽出し、Tweetごとに推定言語表現と降雨強度レベルを紐づける。続いて、これらを推定言語表現別に2年間分集約し、最大件数のレベル値を当該推定言語表現のレベル値とする。

表10 本モデルの月別適用結果(2017)
Table 10 The result of applying the estimation model (2017).

対象月	Tweet数 (件)	一致率 月間平均 (%)	Category毎の 全体に占める 割合(%)		
			1	2	3
1月	409	98.8	94	4	2
2月	394	97.0	92	7	1
3月	704	92.1	87	11	2
4月	968	91.4	83	13	4
5月	679	83.0	61	21	18
6月	763	78.2	32	35	33
7月	990	72.6	20	24	56
8月	1171	73.4	22	27	51
9月	807	82.5	40	22	38
10月	1658	87.0	74	12	14
11月	447	91.5	87	8	5
12月	223	97.2	93	7	0
平均		87.1	---	---	---

表11 一致率年間平均パターン別比較※(2016)
Table 11 Annual average of matched rates by pattern※(2016).

(%)					
A	B	C	D	E	F
86.7	80.4	88.7	86.1	90.5	87.7

※XRAINの週及時間は5分

次に、実験2の手順を述べる。表6の結果をCategory別にせず算出した年間一致率の結果を表11に示す。この結果、パターンEの一致率が最も高い。したがって、全ての推定言語表現について、レベル値をパターンEにより設定する。

各実験におけるモデルを2017年の有効Tweetに適用し、得られた推定値を実際のXRAINと照合した。この結果を本モデルによる結果と併せて表12に示す。

表12のとおり、本モデルの設定方法が最も高精度であった。以降、各実験結果を考察する。

実験1では、Categoryごとに結果は異なった。まず、Category1の推定言語表現は設定値が本モデルと変わらず、結果として一致率も同一となった。しかし、Category2及びCategory3ではレベル値を年間一つに集約すると季節的な降雨変動が反映されず、冬期または夏期を中心に実際のXRAINと不一致となる事例が大幅に増えた。したがって、月別に分割して設定する手法は有効と考えられる。

実験2では、推定値設定パターンがCからEに変更されたCategory3の推定言語表現のみ、本モデルとの違いが生じた。したがって、Category3の推定言

表 12 モデルの条件変更による一致率の比較 (2017)
Table 12 Comparison of rates with model condition change (2017).

(%)			
対象月	実験 1	実験 2	本モデル
1 月	89.1	98.0	98.8
2 月	90.4	96.2	97.0
3 月	85.2	91.8	92.1
4 月	83.1	89.2	91.4
5 月	79.5	83.0	83.0
6 月	57.4	68.5	78.2
7 月	49.2	60.1	72.6
8 月	50.6	62.7	73.4
9 月	61.6	79.5	82.5
10 月	63.2	72.9	87.0
11 月	88.1	91.4	91.5
12 月	89.4	96.2	97.2
年間平均	73.9	82.5	87.1

語表現が多く使われる月 (6~8 月) における一致率が低くなった。Category3 の推定言語表現は、降雨の時間的な変動幅が大きいゆえに統計的精度が低い。この時間的な変動を捉える手法として、頻出値を採用する手法 (パターン E) は平均値を採用する手法 (パターン C) に比べやや精度が落ちた。したがって、Category ごとに推定値設定パターンを別々に割り当てる手法は有効と考えられる。

以上の結果から、本モデルの構成要素の「月」及び「Category 分割」の必要性が示され、本提案手法は「SNS における降雨に関する言語表現から降雨強度レベルを高精度で推定する手法」として有効と結論付けることが可能と考える。

5.3 精度に関する考察

以上の結果のとおり、本推定値は物理情報である XRAIN と高精度で一致する結果を得た。この精度を一般的な公開気象情報と比較すると、本研究の局地的かつリアルタイム配信の条件に最も近い情報として気象庁が公開する「高解像度降水ナウキャスト」(250m メッシュ・5 分ごと 1 時間までの予測) がある。当該精度はしきい値が大きいほど予測精度が低くなるが、しきい値 1 mm h^{-1} における 5 分後のスレットスコア^(注13) のスコア値は最も高い場合で約 0.8 弱とされる [24]。このため、本研究における推定値の時間的優位性を差し引いても、同予測精度とほぼ同等の精度で

あると言える。したがって、本研究での SNS の言語表現に基づく推定値は一般的な気象情報として有意性があると考えられる。

一方、防災上の数値は一般的に大雨の予測に関する適中率により示されるが、気象庁では、局地モデル (LFM)^(注14) を中心として解析される。しかし、当該モデルの精度はしきい値が 1 mm h^{-1} 以上かつ検証格子 10 km の条件下でのスレットスコアのスコア値として約 0.4 前後とされ [25]、現状では高い精度の予測は難しい。当該モデルは予測値であり本研究とは時間幅及び対象面積も異なるため単純比較はできない。しかし、現状の物理的観測情報以外に本研究による推定値を補助的な情報として組み合わせることは、総合的な精度向上のための選択肢の一つと考えられる。

6. む す び

本研究では、Tweet を用いて言語表現から降雨強度レベルを推定する手法を提案した。推定元の Tweet は、先行研究の利用により推定位置情報を付加して空間的精度が高い Tweet を拡大し、独自の抽出アルゴリズムにより時間的精度を高めてノイズを排除した。また、言語表現から降雨強度レベルを推定するモデルは、推定言語表現を統計的精度に応じて分類し、当該分類ごとに推定値設定パターンを決定して推定言語表現ごと・月ごとに推定値を割り当てる手法で構築した。そして、適用事例では実際の物理情報と高精度で一致する結果を得た。

本提案手法は地域ごとに推定言語表現を抽出し、当該地域の XRAIN との照合結果から推定値を得る手法とした。これは、地域ごとに言語表現 (方言等)、気候差、地形、気候に対する感受性、人口密度の分布により、異なる影響を与えることを勘案したことによる。したがって、本論文では関東地域の事例を紹介したが、本手法を他地域に適用し、当該地域の実態に基づくレベル値を設定することで、各地域において独自に降雨強度を推定することが可能になると考える。

本提案手法は、投稿情報の利用による天気情報の表示の応用や、物理情報の観測対象地域における代替

(注13)：スレットスコア：予報、または実況で「現象あり」の場合の予報適中事例数に着目して予報精度を評価する指標。気象庁「数値予報研修テキストで用いた表記と統計的な指標」<https://www.jma.go.jp/jma/kishou/books/nwptext/42/Appendix-B.pdf>

(注14)：局地モデル (LFM)：防災気象情報作成支援、航空予報支援を主な目的とする水平 2 km メッシュでの高解像度数値予報モデル (毎時データ)。発達した積乱雲を直接表現することができ、集中豪雨や突風など局地的な現象を主な予測対象とする。気象庁予報部数値予報課「局地モデル LFM のモデルの特性と利用上の留意点について」<https://www.jma.go.jp/jma/kishou/minkan/koushu131120/shiryou1.pdf>

値としての利用も期待できる。このように本研究の成果は、言語表現から物理情報を推定する手法の一つとして有効であると考えられる。

謝辞 本研究は JST (科学技術振興機構) 戦略的創造研究推進事業 CREST 「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」、文部科学省研究委託事業「地球環境情報統合プログラム (DIAS-P)」及び「地球環境情報プラットフォーム構築推進プログラム (DIAS-PF)」の支援を受けたものである。また、Tweet データ及び関連プロダクトの使用を許諾下さった東京大学の豊田正史教授、吉永直樹准教授には御礼申し上げます。

文 献

- [1] 佐野仁美, 生駒栄司, 喜連川優, “X バンド MP レーダデータの可視化および利用環境の構築,” 情報処理第 77 回全国大会, 第 77 回全国大会講演論文集, 2015(1), pp.475–476, March 2015.
- [2] H. Sano, E. Ikoma, M. Kitsuregawa, and M. Oguchi, “Implementation of disaster prevention-oriented information service platform of XRAIN on DIAS,” 2017 IEEE 6th International Congress on Big Data, pp.398–405, June 2017.
- [3] H. Sano, E. Ikoma, M. Kitsuregawa, and M. Oguchi, “Proposed method for disaster prevention-oriented synthetic rainfall of C-band radar and XRAIN,” Proc. 12th ACM International Conference on Ubiquitous Information Management and Communication (IMCOM2018), 9-1, Jan. 2018.
- [4] 榑 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.
- [5] A. Musaeu and C. Pu, “Landslide information service based on composition of physical and social sensors,” 2017 IEEE 33rd International Conference on Data Engineering (ICDE), pp.1415–1416, April 2017.
- [6] T. Sakaki, M. Okazaki, and Y. Matsuo, “Earthquake shakes Twitter users: Real-time event detection by social sensors,” 19th International Conference on World Wide Web, pp.851–860, 2010.
- [7] 長野伸一, 上野晃嗣, 長 健太, “ソーシャルセンサからの鉄道運行情報検出システムの開発,” 信学論 (D), vol.J96-D, no.10, pp.2262–2273, Oct. 2013.
- [8] C. Maru, M. Enoki, A. Nakao, S. Yamamoto, S. Yamaguchi, and M. Oguchi, “Development of failure detection system for network control using collective intelligence of social networking service in large-scale disaster,” 27th ACM Conference on Hypertext and Social Media (HT2016), pp.267–272, July 2016.
- [9] 馬賀嵩士, 三川健太, 後藤正幸, 吉開朋弘, “気象情報と Tweet データの統合的分析による体感気温の定量化とその需要予測への応用,” 信学論 (D), vol.J101-D, no.7, pp.1037–1051, July 2018.
- [10] 国土交通省, “土砂災害の前兆現象把握に「つぶやき情報」活用～官民共同研究 (平成 26～27 年度) の開始～,” July 2014, <http://www.nilim.go.jp/lab/bcg/kisya/journal/kisya20140714.pdf>.
- [11] 森田直志, 早川 誠, 高尾典佑, “ソーシャルメディア情報を活用した防災システム,” 富士通, vol.68, no.6, pp.62–67, Nov. 2017.
- [12] 武田邦敬, 滝口茂隆, 高橋哲朗, 山影 譲, 渡部 勇, “豪雨時の Twitter データを活用した災害事象の検知,” 砂防学会研究発表会概要集, 2013, B, pp.218–219, May 2013.
- [13] Y. Tzeng, J. Jiang, and P. Cheng, “Event duration detection on microblogging,” 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, vol.1, pp.16–23, 2012.
- [14] L. Butgereit, “Crowdsourced weather reports: An implementation of the μ model for spotting weather information in Twitter,” 2014 IST-Africa Conference Proceedings, pp.1–9, 2014.
- [15] S. Kotozaki, K. Tamura, and H. Kitakami, “Identifying burstiness of local topic using location-based burst detection with a classifier technique,” 2014 IEEE 7th International Workshop on Computational Intelligence and Applications (IWCIA), pp.225–230, 2014.
- [16] T. Sakai and K. Tamura, “Identifying bursty areas of emergency topics in geotagged tweets using density-based spatiotemporal clustering algorithm,” 2014 IEEE 7th International Workshop on Computational Intelligence and Applications (IWCIA), pp.95–100, 2014.
- [17] 国友 優, 神山嬢子, “Twitter を活用した土砂災害の前兆・発生状況把握の可能性,” 土木技術の総合情報誌, vol.57, no.9, pp.18–21, Sept. 2015.
- [18] K.K. Win, K. Zettsu, and K. Sugiura, “Geovisualization and correlation analysis between geotagged Twitter and JMA rainfall data: Case of heavy rain disaster in Hiroshima,” 2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM), pp.71–76, July 2015.
- [19] T. Funayama, Y. Yamamoto, M. Tomita, Y. Kajita, S. Tajima, K. Utsu, and O. Uchida, “Disaster mitigation support system using web services and SNS information,” 2015 13th International Conference on ICT and Knowledge Engineering (ICT & Knowledge Engineering 2015), pp.42–45, 2015.
- [20] 影澤秀明, 廣井 慧, 奥矢 淳, 香取啓志, 加藤 朗, 砂原秀樹, “Twitter を用いたセンシングシステムの提案と考察,” マルチメディア, 分散協調とモバイルシンポジウム 2014 論文集, 2014, pp.725–732, July 2014.
- [21] S.E. Middleton, L. Middleton, and S. Modafferi, “Real-time crisis mapping of natural disasters using social media,” IEEE Intelligent Systems, vol.29, no.2, March–April 2014.
- [22] M. Itoh, N. Yoshinaga, and M. Toyoda, “Word-

clouds in the sky: multi-layer spatio-temporal event visualization from a geo-parsed microblog stream,” 20th International Conference on Information Visualization (IV2016), pp.282-289, July 2016.

- [23] 伊藤正彦, 吉永直樹, 豊田正史, “マイクロブログストリーム中の位置参照表現に着目した実世界イベントの時空間可視化,” 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), Feb.-March 2016.
- [24] 加藤亮平, 清水慎吾, 下瀬健一, 前坂 剛, 岩波 越, 中垣 壽, “局地的短時間強雨に対する気象庁高解像度降水ナウキャストの予測精度検証,” 第17回非静力学モデルに関するワークショップ, pp.55-56, Dec. 2015.
- [25] 気象庁: 数値予報解説資料 (数値予報研修テキスト) [第51巻 (平成30年度) 第10世代数値解析予報システムと数値予報の基礎知識] 第1部 最新の数値予報 第3章: 検証・事例調査, <https://www.jma.go.jp/jma/kishou/books/nwptext/51/1.chapter3.pdf>

(2019年1月25日受付, 4月6日再受付,
5月22日早期公開)



佐野 仁美

1994 お茶の水女子大学・理学部卒。2016 同大学院博士課程 (後期) 在学中。日本銀行, 中央青山監査法人, お茶の水女子大学 (非常勤講師) を経て, 現在, 東京大学地球観測データ統融合研究機構に特任研究員として従事。情報処理学会会員。



伊藤 正彦

2007 北海道大学大学院情報科学研究科博士後期課程了。博士 (情報科学)。同年, 同大学院博士研究員。2009 東京大学生産技術研究所助教を経て, 2014 年より同大学特任准教授, 及び, NICT 主任研究員。2019 年 4 月より北海道情報大学准教授。情報可視化, ユーザインタフェースの研究に従事。IEEE CS, 日本データベース学会, 情報処理学会, 芸術科学会各会員。



生駒 栄司 (正員)

2000 年東京大学大学院工学系研究科電子情報工学専攻博士課程了, 博士 (工学)。同大学空間情報科学研究センター助手・助教, 地球観測データ統融合研究機構特任助教を経て, 2010 年より同特任准教授。現在, データベース工学分野の研究に従事。電子情報通信学会, 日本データベース学会, 情報処理学会各会員。



喜連川 優 (正員:フェロー)

1983 年東京大学工学系研究科情報工学専攻博士課程修了, 工学博士。国立情報学研究所所長, 東京大学生産技術研究所教授。情報処理学会会長 (2013-2015 年), 日本学術会議情報学委員長 (2014-2016 年) などを務める。データベース工学の研究に従事。ACM SIGMOD E. F. Codd Innovations Award, 電子情報通信学会業績賞/功績賞, 情報処理学会功績賞, 全国発明表彰「21 世紀発明賞」, C&C 賞などを受賞。2013 年には紫綬褒章, 2016 年にはレジオン・ドヌール勲章を受賞。ACM, IEEE, 電子情報通信学会, 情報処理学会フェロー。



小口 正人 (正員)

1990 慶應義塾大学・理工学部卒。1995 東京大学大学院博士課程了。博士 (工学)。学術情報センター中核的研究機関研究員, 東京大学生産技術研究所特別研究員, 中央大学研究開発機構助教授, お茶の水女子大学助教授を経て, 現在, 同教授。ネットワーキングコンピューティング・ミドルウェアに関する研究に従事。IEEE, ACM, 電子情報通信学会, 情報処理学会, 日本データベース学会各会員。