ツイートを用いた生物季節観測の見頃推定手法による情報提供の検討

遠藤 雅樹[†,††], 三富 恵佑[††], 佐伯 圭介[††], 江原 遥[†††], 廣田 雅春[††††], 大野 成義[†], 石川 博[††]

[†]職業能力開発総合大学校 基盤ものづくり系, [††]首都大学東京大学院 システムデザイン研究科
[†††]産業技術総合研究所 人工知能研究センター, [††††]大分工業高等専門学校 情報工学科

【あらまし】近年，通信ネットワークやスマートフォン・タブレットの性能向上や普及に伴い，Web上では大量の情報がリアルタイムに発信されている．我々は，近年急速に普及し注目されているWebサービスの１つであるTwitterに着目し，Twitterで発信される大量の情報の中から位置情報付きツイートの分析により，実世界のリアルタイムな「今」の状況を捉え，季節に応じて話題となる生物の見頃を推定する手法を検討した．本手法により，生物季節変化の開始・見頃・終了を推定することで，日本国内の各地域において「今」見頃となっている生物の情報が取得可能になれば，それぞれの地域の季節変化に応じた情報提供を行える可能性がある．本稿では，2015年桜と2015年紅葉を実験対象に，気象庁が全国の気象官署で統一基準により観測している生物季節観測の実データと比較することで，提案手法の有効性を確認する．

【キーワード】傾向推定，生物季節観測，Twitter

Study of information provided by the best time to see estimation method of phenological observations using tweets

Masaki Endo [1), 2)], Keisuke Mitomi [2)], Keisuke Saeki [2)], Yo Ehara [3)], Masaharu Hirota [4)], Shigeyoshi Ohno [1)], Hiroshi Ishikawa [2)]

1) Division of Core Manufacturing, Polytechnic University
2) Graduate School of System Design, Tokyo Metropolitan University
3) Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology
4) Department of Information Engineering, National Institute of Technology, Oita College

【Abstract】 In recent years, performance improvement and dissemination of communication networks and smartphone and tablets, a large amount of information is transmitted in real time on the Web. We consider a method to estimate a best time of phenological that becomes a topic depending on the season capture a real-time situation of the real world to analyze position information with Tweets. In this paper, we considered effectiveness of the proposed method compared to an actual data of phenological observations of the Japan Meteorological Agency as an experimental subject cherry blossoms in 2015 and autumn leaves in 2015.

【Keywords】 Trend Estimate, Phenological Observation, Twitter

1.はじめに

近年，通信ネットワークやスマートフォン・タブレットなどのデバイスの急速な性能向上や普及に伴い，多種多様で膨大なデジタルデータが生成され，Web上に流通・蓄積されるようになった．その中でも，ソーシャルネットワーキングサービス（SNS）と称されるWebサービスが急速に普及し注目を集めている．SNSの個人での利用は，総務省の平成25年通信利用動向調査[1]によると，13歳～59歳では5割を超え利用が拡大している．マイクロブログを提供するSNSとして代表的であるTwitter[2]は，リアルタイムなコミュニケーションツールとして活用できることから日本国内での利用者も多く日々大量の情報発信が行われている．

SNSを通して利用者が発信する大量の情報の中から

実世界のリアルタイムな「今」の状況を捉えることは，有用なSNSの活用法であり，その状況に対応したリアルタイムな「今」の情報提供が可能である．

ここで，Webを利用した観光情報提供について述べる．地域観光情報は，SNSが普及する以前から，自治体・観光協会・旅行会社などが中心となってWebページを通じて提供してきた．さらに，近年のTwitterやFacebookなどのSNSの普及に伴い情報提供が容易になったことから，詳細な情報を観光スポットごとに情報発信する取り組みも増加している．しかし，適時性や話題性を持つ観光情報の提供を行うには，更新頻度が増加するなど，情報提供側のコスト増につながる．そのため，地域や観光スポットごとの取り組みには差異があり，安定した情報提供ができる地域や観光スポットは少ない現状である．

経済産業省のWebを通して提供される観光情報についての実態調査[3]においても，地域の旬でリアルタイムな情報や地元ならではの情報提供を期待するものの，更新頻度が低いことが多くガイドブックなどの出版物と同程度の情報しか提供されていない．また，各自治体・観光協会・旅行会社が独立してWeb情報を公開しており，旅行先地域単位でまとまった情報提供が行われているものの，それぞれの観光スポットの「今」を取得するためには，情報収集にコストがかかるなどの意見が挙げられている．

旅行者は，よりリアルタイムな観光情報提供を望んでいる．そのため，観光地域の時期や時間帯に応じて，刻々と変化する旅行者に有用な実世界の「今」の情報提供が必要であると考えられる．ここで，実世界の「今」とは，花が咲いて見頃である状況や訪問地域で開催されているイベント情報，局所的な大雨などの防災情報など，旅行者が情報を取得する時点での季節やイベント，防災を検討する上で重要な要因となるものを想定している．

我々は，実世界の「今」と想定した状況の中から，観光として行う花見や紅葉狩りなどの生物季節観測に着目し，各地域での見頃推定を行う手法について検討する．観光情報として必要な情報は，花が咲く前や散った後のピーク前後ではなく見頃となるピーク期を推定する必要がある．また，地域や場所によって見頃は異なるため地域や場所ごとに見頃を推定する必要もある．

この生物季節観測の見頃推定を行う上で必要となるリアルタイム性を持つ情報取得のためには，即時性のある情報を多く収集する必要がある．本研究は日本国内の利用者が多いSNSであるTwitterを利用することとした．

本稿では，日本国内で投稿されるツイートを利用した生物季節観測の見頃推定を行う手法を提案する．そして，提案手法を用いて行った2014年の紅葉と2015年の桜での実験により，地域別・観光スポット別での見頃推定実験を行った結果を報告する．以降では，2章に関連研究，3章で提案手法を示し，4章で実験方法と実験結果を示す．最後に，5章で総括する．

2.関連研究

SNSなどの普及に伴い今後もデジタルデータの量は飛躍的に増大することが予想され，この大量のデジタルデータの有効活用に関連した研究が数多く行われている．マイクロブログに関しても，発信された情報を分析しマイクロブログ内での動向から実世界の動向の把握や予測に結び付ける研究が様々な分野で行われている[4]．Kleinberg[5]は，時系列データにおいてキーワードが急激に増加する現象である「バースト」を検出できることを示している．このバースト検知を用いてトレンドを分析する研究も行われている．落合ら[6]は，マイクロブログを対象として，場所に特有の季節変動などに依存しない静的特徴語と場所を含む期間ごとに変化するトピックである動的特徴語を利用した同名地名の曖昧性解消手法を提案している．中嶋ら[7]は，旅行者の発信した位置情報付きツイートの特徴から「食事」，「景観」，「行動」の

3つに分類した情報を用いて好みに合わせた観光ルートの推薦手法を提案している. 倉田ら[8]は, 位置情報が付加されたツイートから, 実空間上のイベントを検知するシステムを構築している. 各時間軸帯での頻出単語上位10件を抽出することで, ある時間のある場所でどのようなイベントが盛り上がっているかを知ることができる. また, 榊ら[9]は, Twitterからリアルタイムなイベントを推定する研究として地震や台風などのイベントを検出する手法を提案している.

このように, イベントを抽出する手法や場所を取得する手法については多くの議論がなされているが, ツイートから生物季節観測の開始と終了を推定し, ピーク期である見頃を取得する手法については新たな議論である. 我々は, 時系列データ処理の手法として, 気象データ分析[10]や株式などの取引市場におけるテクニカル分析[11]において, 一般的に用いられる移動平均を利用した. これを, Twitterの時系列データに対し適用した生物の見頃推定手法について提案する.

3.観光情報の分析に関する研究

本章では, 日本国内の季節変化をTwitterから取得するための分析対象データ収集と見頃推定を行う手法について記述する. 図1に提案手法についての概要を示す.

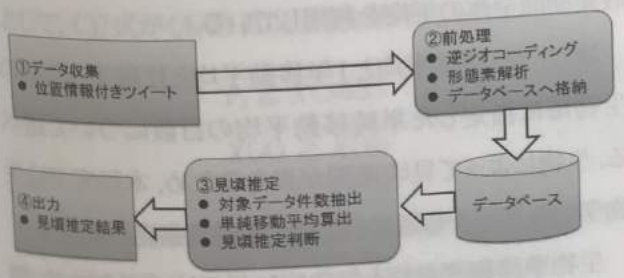


図1 システム概要

3.1.データ収集

本節では, 図1に示した①データ収集の手法について記述する. Twitterから発信された位置情報付きツイートの中で, 日本の領土を含む範囲である緯度経度が10進法表記で120.0≦経度≦154.0かつ20.0≦緯度≦47.0である位置情報付きツイートを収集対象とし, 位置情報付きツイートのデータの収集には, Twitter社が提供するAPIの1つであるStreaming API[12]を用いて収集した.

次に, 収集したデータ数について述べる. 橋本ら[13]の研究によると, 日本国内で発信されるツイートの中で位置情報が付いている割合は約0.18%とツイート全体では非常に少ないデータ数である. しかし, 収集した位置情報付きツイートは, 表1に示す推移例のとおり, 平日でも約7万件, 土日には10万件を超える日もある. 本研究において収集した位置情報付きツイートは, 2015/2/17から2015/12/31までの期間で約2,100万件である. また, 期間中の1日当たりの収集件数は約67,000件であった. このデータセットを用いて次節以降で述べる処理により見頃推定を行った.

表1 位置情報付きツイートの推移例(2015/5/9-6/3)

日付(曜日)	件数	日付(曜日)	件数
5/9(土)	117,253	5/22(金)	92,237
5/10(日)	128,654	5/23(土)	55,590
5/11(月)	91,795	5/24(日)	72,243
5/12(火)	87,354	5/25(月)	82,375
5/13(水)	67,016	5/26(火)	83,851
5/14(木)	88,994	5/27(水)	83,825
5/15(金)	89,210	5/28(木)	85,024
5/16(土)	116,600	5/29(金)	121,582
5/17(日)	126,705	5/30(土)	119,387
5/18(月)	89,342	5/31(日)	81,431
5/19(火)	83,695	6/1(月)	76,364
5/20(水)	87,927	6/2(火)	76,699
5/21(木)	86,164	6/3(水)	78,329

3.2.前処理

本節では図1に示した②前処理について記述する. 3.1節の処理により収集したデータに対して, 逆ジオコーディング・形態素解析・データベースへの格納の処理を行う.

逆ジオコーディングは, 収集した個々のデータ(ツイート)の緯度経度情報から都道府県・市区町村・町名を特定した. この処理には, 独立行政法人農業・食品産業技術総合研究機構の簡易逆ジオコーディングサービス[14]を用いた. 例として, (緯度, 経度)

＝(35.7384446, 139.460910)を逆ジオコーディングすると，東京都・小平市・小川西町二丁目が得られる．

形態素解析は，収集した個々のデータ(ツイート)の本文を形態素解析器「Mecab」[15]を用いて，分かち書きを行う．例として，「桜がきれいです。」は，「桜/名詞，が/助詞，きれい/名詞，です/助動詞，。/記号」と分割される．

データベースへの格納は，データ収集・逆ジオコーディング・形態素解析の処理を行った結果から見頃推定に必要となるデータをデータベースに格納する．本研究において利用したデータは，ツイートID・ツイート投稿時刻・ツイート本文・形態素解析結果・緯度・経度である．

3.3.見頃推定

本節では図1に示した③見頃推定について記述する．見頃推定は，対象データ件数抽出・単純移動平均算出・見頃推定判断の処理を行う．

対象データ件数抽出は，見頃推定に関連する語を含むデータ(ツイート)をデータベースから抽出し，分析単位とした日毎にデータ件数を集計する．ここで，見頃推定に関連する語は，次の2種類とした．1つ目は，対象語と定義し，表2に示した生物名や季節変化を表す漢字・ひらがな・カタカナ表記を含む語とする．2つ目は，共起語と定義し，表2に示した対象語と共起する観光スポット名とする．

表2 対象語一覧

項目	対象語
さくら	さくら，桜，サクラ
かえで	かえで，楓，カエデ
いちょう	いちょう，銀杏，イチョウ
こうよう	こうよう，もみじ，紅葉，黄葉，コウヨウ，モミジ

次に，単純移動平均算出について述べる．見頃判断の基準に移動平均を用いて，前述の対象データ件数抽出により日毎に集計したデータ件数を用いて単純移動平均を算出する．単純移動平均の日数についての概要を図2に示す．見頃推定日の前日から過去に遡ったデータ数を用いて(1)式を用いて単純移動平均を求める．

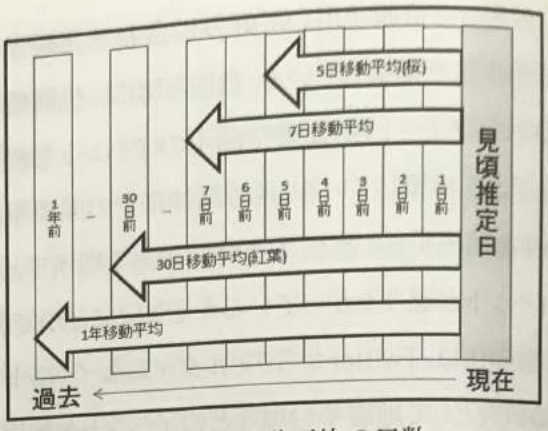


図2 単純移動平均の日数

$$X(Y)=\frac{P_1+P_2+\cdots+P_Y}{Y} \quad \cdots(1)$$

$X(Y)$：Y 日移動平均

P_n：n 日前のデータ数

Y：算出対象期間

単純移動平均の基準とする日数については，7日移動平均と1年移動平均とする．7日移動平均は，前述の表1で示したとおり，位置情報付きツイートの推移が平日に比べ土日に増加する傾向があることから，1週間毎(7日)の周期を見頃推定の基準とした．また，生物季節観測は桜や紅葉に代表されるように1年毎の事象が多いことから，過去1年間の推移である1年移動平均を見頃推定の基準とした．ただし，4章で述べる実験では，データ収集が過去1年分を取得できていないため暫定的にデータ収集期間全体の平均を利用している．

次に，7日移動平均と1年移動平均と比較するための生物毎に指定した単純移動平均の日数について述べる．生物によって見頃期間が異なるため，本研究では生物学的な期間から個々の生物の日数を指定した．

生物季節観測を行う気象庁[16]では，「さくら」は，開花日と満開日の2項目が観測対象である．「さくらの開花日」[17]とは，標本木に5～6輪以上の花が咲いた最初の日を指す．「さくらの満開日」は，標本木で約80%以上のつぼみが開いた状態となった最初の日としている．また，「さくら」は，一般的に開花から満開になるまでの日数が，

5日程度である．よって，本研究において「さくら」は，5日移動平均を基準とした．

次に，「かえで」は，紅葉日と落葉日の2項目が気象庁の観測対象である．「かえでの紅葉日」[18]とは，標本木全体を眺めたときに，大部分の葉の色が紅色に変わった最初の日を指す．「かえでの落葉日」は，約80%の葉が落葉した最初の日としている．

また，「いちょう」は，黄葉日と落葉日の2項目が気象庁の観測対象である．「いちょうの黄葉日」[19]とは，標本木全体を眺めたときに，大部分の葉が黄色に変わった最初の日を指す．「いちょう」の落葉日は，約80%の葉が落葉した最初の日としている．

気象庁の観測する「かえで」・「いちょう」の紅葉（黄葉）日から落葉日までの期間は，一般的に1ヵ月（30日）であることから，本研究では表2の「かえで」・「いちょう」・「こうよう」については，30日移動平均を基準とした．

次に，見頃推定判断について述べる．前述の単純移動平均算出により求めた7日移動平均・1年移動平均・生物別移動平均を用いて見頃推定判断を行う．見頃推定判断の条件として2つの条件を指定した．条件1は，(2)式で表すとおり，1日前のデータ数が見頃推定日の単純移動平均以上である．また，条件2は，7日移動平均と生物別移動平均で短い方の日数をA日，長い方の日数をB日として，(3)式が$(A/2)$日以上続いた日とした．

$$P_i \geqq X(365) \quad \cdots (2)$$

$$X(A) \geqq X(B) \quad \cdots (3)$$

本研究の見頃推定判断では，見頃推定日に前述の条件1と条件2が共に成り立つ日を見頃と判断した．

3.4.出力

本節では図1に示した④出力について記述する．出力は，前節までの処理によって見頃推定を行った結果を利用した可視化を想定している．本稿では，横軸に日付，縦軸にデータ数を取った時系列グラフに，見頃推定結果を反映させた可視化を行っている．旅行者に有用な可視化手法の検討については，今後の課題としている．

4.実験方法と実験結果

本章では，3章で述べた提案手法についての見頃推定の実験について述べる．4.1節に見頃推定実験に利用したデータセットを示し，4.2節に2015年桜，4.3節に2015年紅葉の見頃推定実験を示す．

4.1.データセット

本実験で使用したデータセットは，3.1節のデータ収集で述べたStreaming APIを用いて収集した2015/2/17から12/31までの期間の日本国内の緯度経度情報を含む位置情報付きツイートである約2,100万件とした．このデータセットを用いて，4.2節に示す2015年の桜，4.3節に示す2015年の紅葉についての見頃推定実験を行った．

4.2.2015年桜の見頃推定実験

2015年桜の見頃推定実験は，3.3節に示した表2の項目「さくら」の「さくら」・「桜」・「サクラ」を含むツイートを対象語とした．実験対象地域は，「東京都」・「石川県」・「北海道」，各地域において気象庁が観測する標本木がある「千代田区」・「金沢市」・「札幌市」とした．また，共起語を利用した実験には，表2の項目「さくら」を含み各地域の観光スポットである「六義園」・「高尾山」・「兼六園」・「円山公園」を含むツイートを用いて行った．

4.2.1.対象地域での対象語による見頃推定結果

4.2節の対象語「さくら」の見頃推定実験について，対象地域を「東京都」・「石川県」・「北海道」とした実験結果をそれぞれ図3・図4・図5に示す．図中の濃灰の棒グラフはツイート数を表し，薄灰部分は，提案手法により見頃と

判断した期間を表す．また，実線・破線・点線はそれぞれ5日移動平均・7日移動平均・1年移動平均を示す．ただし，本節で示す4/28–5/8の期間は，データ収集処理の仕様変更時の不具合により極端に少ないデータ数となっている．

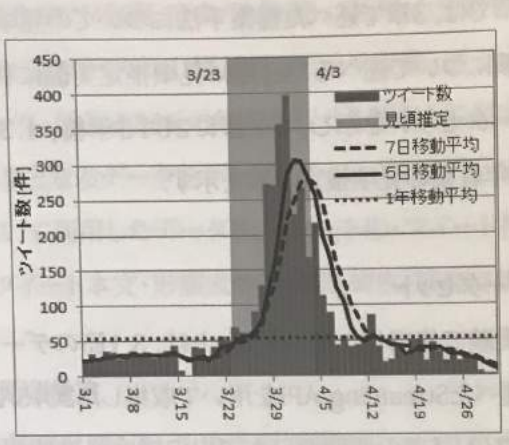


図3 東京都の対象語「さくら」での見頃推定

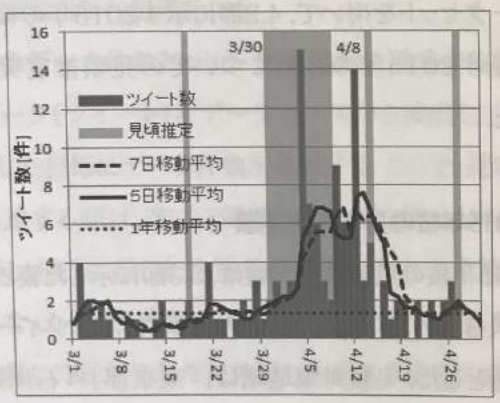


図4 石川県の対象語「さくら」での見頃推定

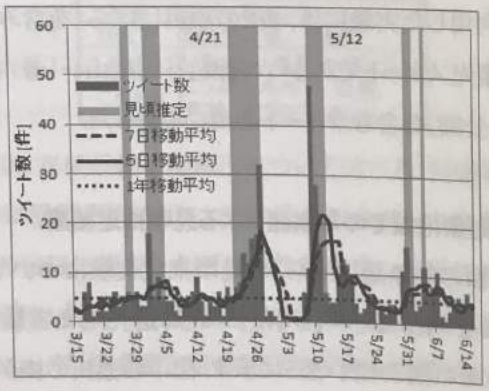


図5 北海道の対象語「さくら」での見頃推定

図3の東京都では，実験対象とした地域で最も多いデータ数が得られ，最多日は約400件となった．提案手法により見頃推定を行った結果，3/23–4/3に見頃推定を確認した．一方で，図4の石川県と図5の北海道では，最多日でそれぞれ15件，約50件と東京都と比較するとデータ数は少ない．しかし，実験の結果，石川県は3/30–4/8，北海道は4/21–5/12頃を，長期間の見頃として推定した．

ただし，データ数が少なくなるほどノイズの影響を受けやすいため，石川県・北海道には，前述の期間の前後にも提案手法により見頃推定が行われた．それらについて確認したところ，生物の「さくら」に関連しないツイートも含まれているが，前部分はつぼみの状態や桜の開花を期待するツイート，後部分は葉桜に関するツイートも確認できた．

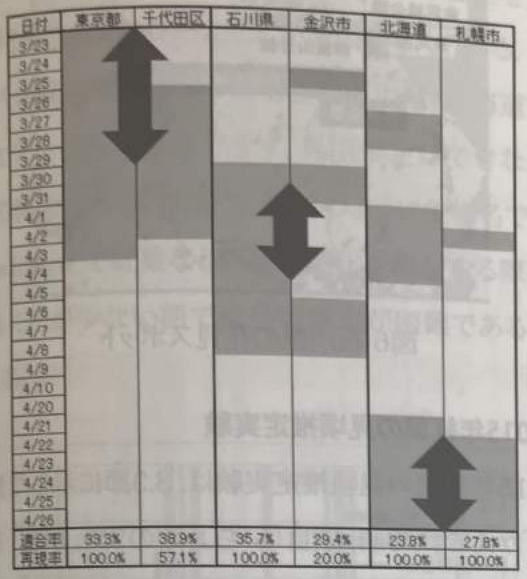
よって，ツイート内容を解析することで，さらに詳細な「さくら」の状態についての見頃推定を行える可能性もある．しかし，本項では，対象語の出現量のみを基準に各地域の見頃を推定する点に着目をしているためツイート内容の解析については言及しない．

4.2.2.対象地域での見頃推定と比較

表3に，対象地域での見頃推定と気象庁の観測データとの比較結果を示す．表内の日付は見頃推定日，各地域の薄灰部は見頃であると判断した日である．例として東京都の見頃は2015/3/23–4/3であり，前項の図3の薄灰部分の見頃推定と同様の日を表している．また，矢印は各地域における気象庁が観測した「さくらの開花日」から「さくらの満開日」までの期間を示す．例として東京都の場合は，千代田区の標本木を基準に観測されており，2015年の「さくらの開花日」・「さくらの満開日」は「3/23」・「3/29」であった．同様に，石川県は金沢市，北海道は札幌市の観測データを利用した．適合率と再現率はそれぞれ(4)式，(5)式により，2015/3/1–6/30の

期間で対象地域ごとに求めた.

表3 対象地域での見頃推定と気象庁の観測データとの比較結果



$$適合率 = \frac{気象庁データと重なった日数}{見頃推定日数} \quad \cdots (4)$$

$$再現率 = \frac{見頃推定と重なった日数}{気象庁の見頃日数} \quad \cdots (5)$$

表3から, 適合率は, 約30%となった. これは, 気象庁の観測する開花から満開との比較のため, 満開から落花までを気象庁の観測期間の矢印に含んでいない影響である. 千代田区の3/30–4/4など, 気象庁の満開観測後の見頃推定は, 満開から落花までの期間の見頃を捉えている. よって, 提案手法により気象庁の観測データを補完することで, 観光に必要となる見頃情報を提供できる可能性を示した. なお, 金沢市のようなデータ数が少ない地域は, 移動平均が極端な変化に弱いため見頃推定に影響がある.

一方, 東京都・石川県・北海道では, 気象庁で観測する標本木の観測データに対して, 各都道県全域のデータを利用したため, 再現率が高くなったと考えられる. 都府県から市区へ地域を絞った場合には, 千代田区・金沢市はデータ数が減少するために再現率が低下している. しかし, 見頃推定の結果は対象地域に限定した情報であり, 観光情報としては地域ごとに情報提供できる可能性を示した. 北海道は, ほとんどのデータが札幌市内のデータであったため, 地域を絞っても再現率は減少しない結果となった.

この実験の結果から, 日毎のツイート数の最多日が約10件と, 最もデータ数が少ない金沢市でも見頃推定を行えることを確認した. よって, 提案手法では最多日に少なくとも10件程度のデータ数を得られる地域の場合は, 見頃推定が可能であると考えられる. ただし, 今回の実験対象地域は, 気象庁が「さくら」を観測する標本木があり, かつ県庁所在地であるなど比較的データ数が多い地域での実験であるため, 他地域ではさらなる検証が必要である.

4.2.3.共起語による見頃推定結果

表4に, 対象語「さくら」と共起する観光スポット名を含むツイートを用いた見頃推定結果を示す. 共起語は, 提案手法により見頃推定ができた観光スポット名である, 「新宿御苑」・「六義園」・「五稜郭」・「兼六園」とした. 表中の数値は, 対象語と共起語を含むデータ数, 薄灰部は提案手法により見頃推定が真となった日を表す. なお, 本実験の検証には, 開花日と満開日の観測のみである気象庁のデータでは各観光スポット別の確認が困難なため, 気象情報会社や公益社団法人などが提供するサービス[20][21]やブログやSNSを利用し, 人手で観光スポットの開花と見頃を確認した. 表4の矢印は, 各観光スポットでの人手で確認した開花から満開までの期間である.

表4から, 各観光スポットのデータ数は非常に少ないが, 観光スポットによる見頃の差異を確認することができる. 特に, 新宿御苑と六義園のように比較的距離が近い場所であっても見頃の差異を確認できる. これは, 提案手法により標本木の観測による気象庁の観測とは異なり, 各観光スポット別の見頃推定を行える可能性を示した. ただし, 一定数以上かつ一定期間継続してツイー

ト数を取得できない観光スポットにおける見頃推定については，提案手法は適用できないため今後の課題とした．

表4 共起する観光スポット名での見頃推定とツイート数

日付	新宿御苑	六義園	五稜郭	兼六園
3/15	0	0	0	0
3/16	0	0	0	0
3/17	0	0	0	0
3/18	1	0	1	0
3/19	0	0	0	0
3/20	0	0	0	0
3/21	1	0	0	0
3/22	0	0	0	0
3/23	0	0	0	0
3/24	3	0	0	0
3/25	0	0	0	0
3/26	0	0	0	0
3/27	0	4	0	0
3/28	0	4	0	0
3/29	3	2	0	0
3/30	5	2	0	0
3/31	1	3	0	0
4/1	4	1	0	0
4/2	2	1	0	0
4/3	0	1	0	0
4/4	2	0	0	2
4/5	1	0	0	2
4/6	0	0	0	0
4/7	0	0	0	0
4/8	0	0	0	0
4/9	0	0	0	1
4/10	0	0	0	0
4/11	0	0	0	1
4/12	1	0	0	2
4/13	0	0	0	0
4/14	0	0	0	1
4/15	0	0	1	0
4/16	0	0	0	0
4/17	2	0	1	0
4/18	4	0	0	0
4/19	1	0	0	1
4/20	0	0	1	0
4/21	1	0	1	0
4/22	0	0	0	0
4/23	0	0	0	0
4/24	0	0	1	0
4/25	0	0	2	1
4/26	0	0	3	0
4/27	0	0	1	0
4/28	0	0	0	0
4/29	0	0	0	0
4/30	0	0	0	0

次に，最もデータ数が少なかった石川県を対象に観光スポット名を共起語とした見頃推定結果について示す．図6に石川県内の花見スポットについて検証を行った結果を示す．その結果，前述の「兼六園」を含む白枠で示した「金沢城公園」・「卯辰山公園」は，見頃推定に成功した．しかし，黒枠で示した「能登町柳田植物公園」・「能登さくら駅（能登鹿島駅）」・「石川県農林総合研究センター」・「芦城公園」は，ツイートから観光スポットとして発見はできたものの，データ数が少ないため見頃推定までは行えない結果となった．よって，データ数の取得にはさらなる議論が必要であることを確認した．

図6 石川県の花見スポット

4.3.2015年紅葉の見頃推定実験

2015年紅葉の見頃推定実験は，3.3節に示した表2の項目「かえで」の3語，項目「いちょう」の3語，項目「こうよう」の6語をそれぞれの項目の対象語とした．実験対象は，前節の「さくら」と同様に，「東京都」・「石川県」・「北海道」・「千代田区」・「金沢市」・「札幌市」とした．また，共起語を利用した実験には，表2の項目「かえで」・「いちょう」・「こうよう」のいずれかの対象語を含み各地域の観光スポットである「六義園」・「高尾山」・「兼六園」・「円山公園」を含むツイートを用いて行った．

4.3.1.対象語での見頃推定

図7に，東京都での対象語「かえで」の見頃推定結果を示す．図中の濃灰の棒グラフはツイート数を表す．薄灰部分は，見頃と判断した期間を表す．また，実線・破線・点線は，それぞれ7日移動平均・30日移動平均・1年移動平均を示す．本節で示す他図も同様の表記とする．なお，12月上旬は，データ収集の不具合によりデータに一部欠損がある．

図7の結果に示すように，対象語「かえで」を含むデータ数は東京都であっても非常に少なく，東京都内でデータ数が最も多い日でも9件であった．見頃推定は，気象庁が観測する「かえでの紅葉日」・「かえでの落葉日」の「12/4」・「12/12」に近い11/26–12/1頃を推定してい

る.また,10月下旬や12月下旬にも見頃と判断している期間もある.これは,高尾山など東京都内では紅葉が早い地域や暖冬の影響で2015年の年末まで紅葉を見ることが可能な地域もあったことが原因と考えられる.しかし,千代田区に地域を絞った場合や石川県・北海道では十分なデータ数が収集できず,見頃推定はできなかった.この結果から,対象語「かえで」のように生物名であってもTwitter上で紅葉などの季節変化を表現する際に使われることが少ない語では見頃推定が困難であると考えられる.

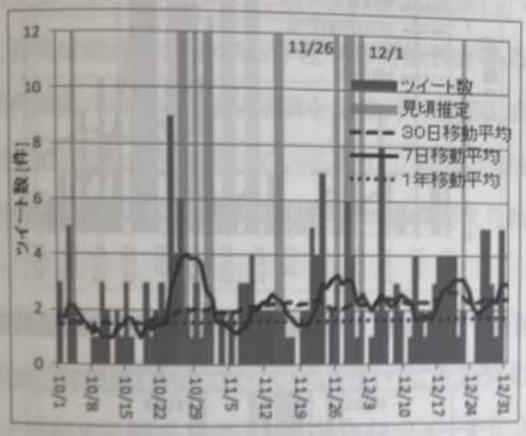


図7 東京都の対象語「かえで」での見頃推定

次に,対象語「いちょう」の見頃推定について,図8・図9・図10に東京都・石川県・北海道,図11・図12に千代田区・札幌市に結果を示す.

図8と図10の東京都・北海道については,最多日に約100件・10件と見頃推定可能なデータ数があり見頃を確認できた.また,図11と図12の千代田区・札幌市は,最多日のデータ数は,6件・約10件であったが見頃の確認ができた.しかし,図9の石川県は,データ数が少なく見頃推定はできなかった.

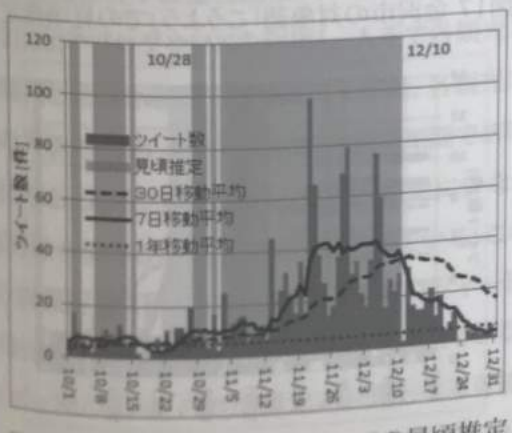


図8 東京都の対象語「いちょう」での見頃推定

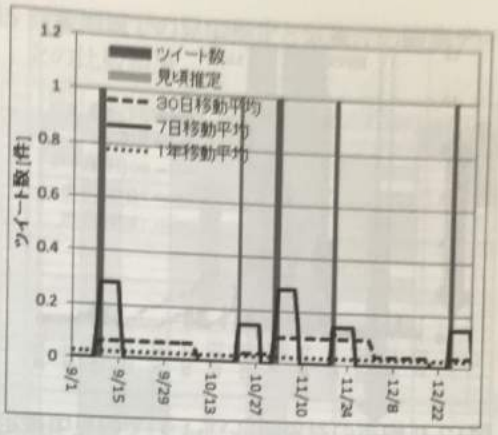


図9 石川県の対象語「いちょう」での見頃推定

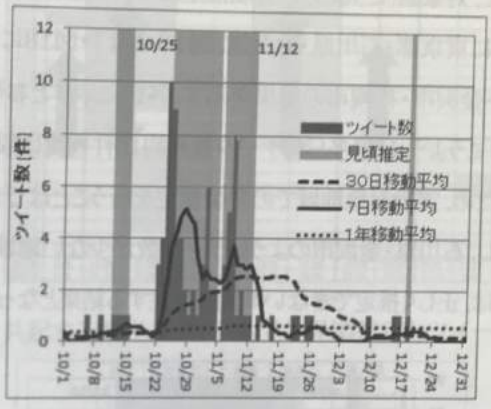


図10 北海道の対象語「いちょう」での見頃推定

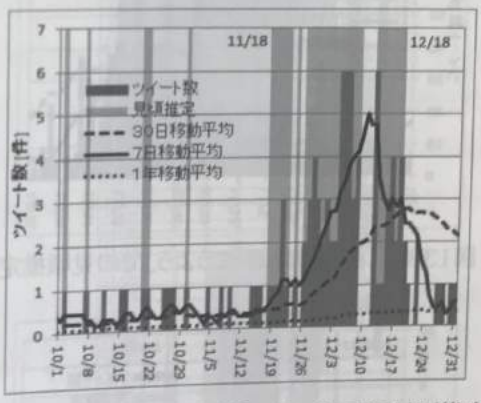


図11 千代田区の対象語「いちょう」での見頃推定

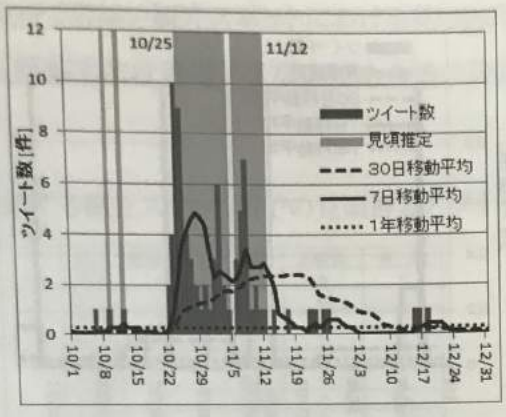


図12 札幌市の対象語「いちょう」での見頃推定

次に，対象語「こうよう」の見頃推定について，図13・図14・図15に東京都・石川県・北海道，図16・図17・図18に千代田区・金沢市・札幌市の結果を示す．各図に示すとおり，対象語「こうよう」は最多日のデータ数が約10件程度は収集できたため，実験対象地域での見頃推定を行うことはできた．しかし，石川県・金沢市のようにデータ数が少ない地域については，正しく推定できない期間も存在する結果となった．

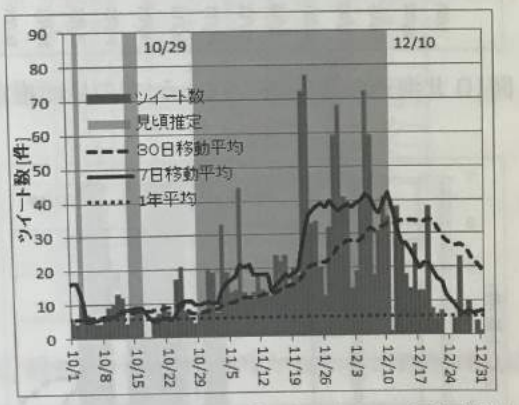


図13 東京都の対象語「こうよう」での見頃推定

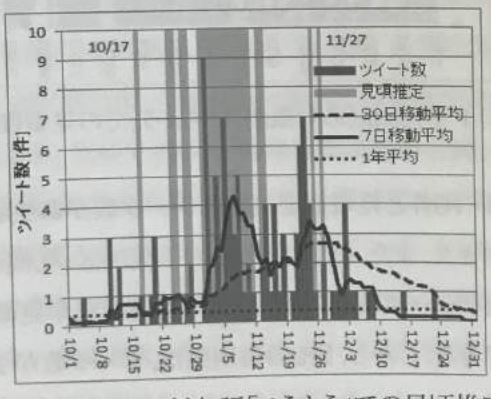


図14 石川県の対象語「こうよう」での見頃推定

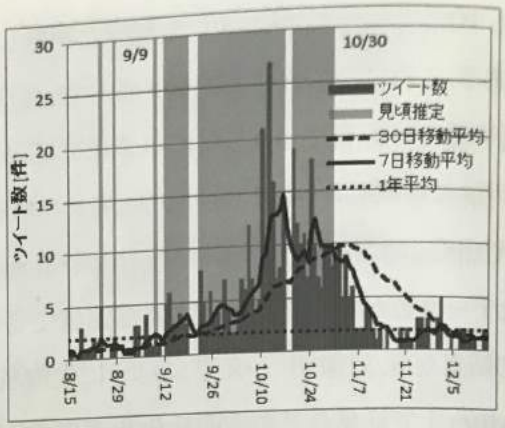


図15 北海道の対象語「こうよう」での見頃推定

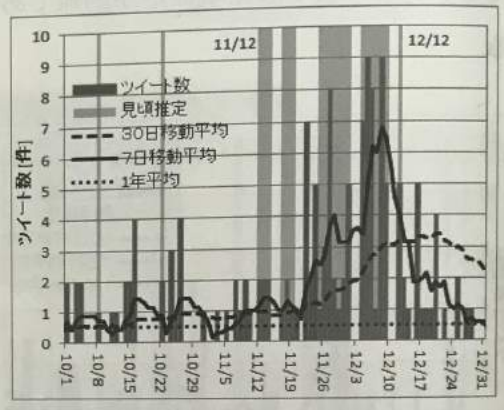


図16 千代田区の対象語「こうよう」での見頃推定

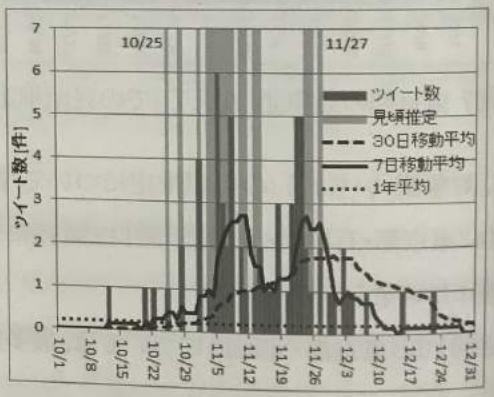


図17 金沢市の対象語「こうよう」での見頃推定

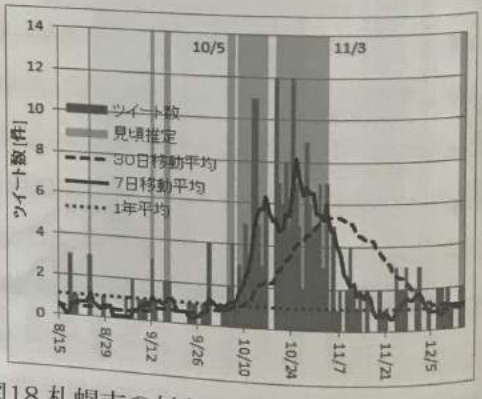


図18 札幌市の対象語「こうよう」での見頃推定

表5に，対象地域での見頃推定と気象庁の観測データとの比較結果を示す．表記は前節で示した表3と同様であり，表内の日付は見頃推定日，各地域の薄灰部は見頃であると判断した日である．ただし，「こうよう」の矢印は，「かえで」と「いちょう」の気象庁の観測を地域ごとに合わせた期間とした．また，適合率と再現率は(4)式と(5)式により，2015/10/1–12/31の期間で対象地域ごとに求めた．

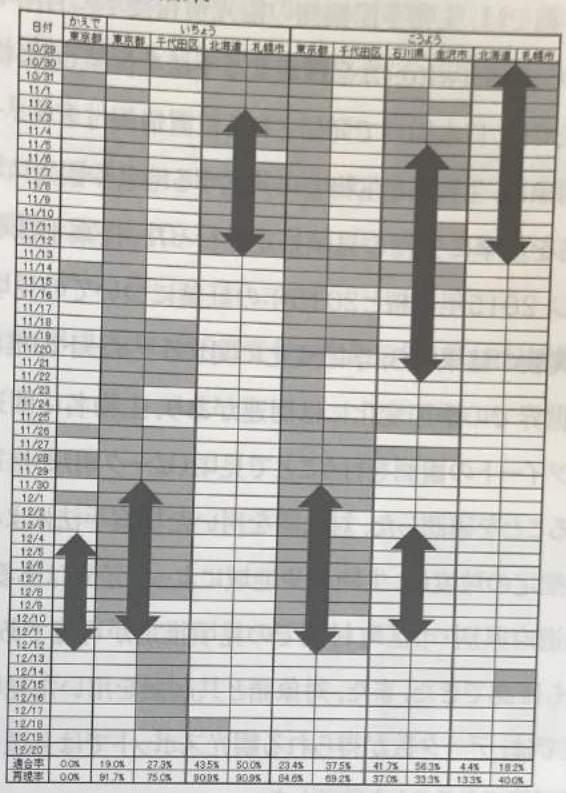
表5から，前節の「さくら」と比較した場合は，データ数が少ないため適合率・再現率は共に低くなった．しかし，気象庁の観測が，大部分の葉が色付いた日である「紅葉(黄葉)日」と80%以上が落葉した最初の日の「落葉日」であるため，見頃は観測日以前にも存在する可能性がある．よって，千代田区の「こうよう」の見頃を気象庁の観測した紅葉日(11/30)以前から推定しているように，提案手法により観光情報に必要な見頃を検出できていると考えられる．

4.3.2.共起語による見頃推定結果

表6に，対象語「かえで」・「いちょう」・「こうよう」のいずれかと共起する観光スポット名を含むツイートによる見頃推定結果を示す．共起語は，提案手法により見頃推定ができた観光スポット名である，「新宿御苑」・「六義園」・「高尾山」・「兼六園」とした．表中の数値は，対象語と共起語を含むデータ数，薄灰部は提案手法により見頃推定が真となった日を表す．なお，本実験の検証には，前述の気象情報会社や公益社団法人などが提供するサービス[22][23]やブログやSNSを利用し，人手で各観光スポットの開花と見頃を確認した．表6の矢印は，各観光スポットでの人手で確認した紅葉(黄葉)の見頃期間である．

表6から，人手で確認を行った見頃期間に少なくとも一部分は含まれる期間を見頃推定できている．このことから，対象観光スポットにおいて一定数のツイートが必要であるものの，人手で観測し情報提供されていない観光スポットについての見頃推定の一助となる可能性を示した．

表5 対象地域での見頃推定と気象庁の観測データとの比較結果



日付	かえで	いちょう				こうよう					
	東京都	東京都	千代田区	北海道	札幌市	東京都	千代田区	石川県	金沢市	北海道	札幌市
10/29											
10/30											
10/31											
11/1											
11/2											
11/3											
11/4											
11/5											
11/6											
11/7											
11/8											
11/9											
11/10											
11/11											
11/12											
11/13											
11/14											
11/15											
11/16											
11/17											
11/18											
11/19											
11/20											
11/21											
11/22											
11/23											
11/24											
11/25											
11/26											
11/27											
11/28											
11/29											
11/30											
12/1											
12/2											
12/3											
12/4											
12/5											
12/6											
12/7											
12/8											
12/9											
12/10											
12/11											
12/12											
12/13											
12/14											
12/15											
12/16											
12/17											
12/18											
12/19											
12/20											
適合率	0.0%	19.0%	27.3%	43.5%	50.0%	23.4%	37.5%	41.7%	56.3%	4.4%	18.2%
再現率	0.0%	91.7%	75.0%	90.9%	90.9%	84.6%	69.2%	37.0%	33.3%	13.3%	40.0%

表6 共起する観光スポット名での見頃推定とツイート数

日付	新宿御苑	六義園	高尾山	兼六園
11/1	0	0	2	0
11/2	0	0	0	0
11/3	1	0	0	0
11/4	0	0	2	1
11/5	0	0	1	5
11/6	0	0	0	0
11/7	1	1	3	1
11/8	0	0	2	2
11/9	0	0	1	0
11/10	1	0	0	1
11/11	1	0	0	0
11/12	1	0	0	0
11/13	1	0	0	1
11/14	1	0	0	1
11/15	0	0	1	1
11/16	0	0	3	1
11/17	0	0	3	1
11/18	0	0	2	1
11/19	0	4	0	0
11/20	1	1	5	1
11/21	0	4	9	3
11/22	2	0	14	2
11/23	1	0	1	4
11/24	0	1	4	0
11/25	0	1	1	0
11/26	0	1	3	0
11/27	0	2	2	2
11/28	4	6	9	0
11/29	2	4	10	0
11/30	1	6	5	0
12/1	0	1	3	0
12/2	0	4	0	0
12/3	0	4	0	1
12/4	0	2	2	0
12/5	4	4	4	0
12/6	5	5	2	0
12/7	3	2	1	0
12/8	0	0	0	0
12/9	1	2	0	0
12/10	1	1	0	0
12/11	0	0	0	0
12/12	2	1	0	0
12/13	7	1	0	0
12/14	0	0	1	0
12/15	1	1	1	0
12/16	1	0	0	0
12/17	0	0	0	0
12/18	0	1	0	0
12/19	0	0	0	0
12/20	0	0	0	0
12/21	0	0	0	0
12/22	0	0	0	0

5.おわりに

本稿では，生物季節観測の観光情報提示に有用な生物の見頃をTwitterから推定する手法を提案した．提案手法では，日本国内で発信される位置情報付きツイートを対象に，生物名と生物名に共起する地名や観光スポット名を基準に生物の見頃推定を行った．提案手法を利用し，2015年の桜と2015年の紅葉についての見頃推定実験の結果から，季節変化に関するツイートの推移と実世界での季節変化には関連があり，生物名に関連するツイートの観測を行うことで見頃（ピーク期）を推定できることを確認した．対象語を用いた提案手法による見頃推定の粒度は，生物名や地域によって差異はあるが，都道府県別・市区町村別での見頃推定が可能であることも確認できた．また，対象語と共起語を用いた見頃推定では，データ数が得られる観光スポットでは見頃推定が可能であることも確認した．

一方で，対象語「かえで」のように，生物名がTwitter上で季節変化を表現する際に使われることが少ない語での見頃推定は困難であるため，対象語の選択は今後さらなる議論が必要である．さらに，都道府県・市区町村・観光スポット名別の見頃推定において，データ不足により推定ができない課題については，Flickrなど他の位置情報付きデータを取得可能なSNSや位置情報を持たないツイートを利用するなど，今後さらなる議論が必要である．

本稿の提案手法を利用することで，位置情報付きツイートを利用した見頃推定により各地域の「今」を捉えリアルタイム性を持つ観光情報提示の可能性が確認できた．今後は提案手法が他の生物にも適応することを検証し，さらに，各地域の「今」を推定することで，旅行者が旅先で開催されているイベント情報や災害情報をリアルタイムに取得するシステムへの拡張を検討していく予定である．

[参考文献]

[1]総務省：平成25年通信利用動向調査の結果，入手先 http://www.soumu.go.jp/johotsusintokei/statistics/data/1406271.pdf, 2014-6.

[2]Twitter：公式サイト，入手先 https://twitter.com/，閲覧（2014-3）.

[3]経済産業省：平成18年度ITを活用した観光情報提供の在り方に関する実態調査，入手先 http://www.meti.go.jp/report/downloadfiles/g70629a01j.pdf, 2006-3.

[4]奥村学：マイクロブログマイニングの現在，信学技報，NLC2011-59, pp.19-24, 2012-2.

[5]J.Kleinberg：Bursty and hierarchical structure in stream, In Proc. of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1-25, 2002.

[6]落合桂一，鳥居大祐：時間変化する特徴語によるマイクロブログ地名曖昧性解消，情報処理学会論文誌データベース，Vol.7, No.2, pp.51-60, 2014-6.

[7]中嶋勇人，新妻弘崇，太田学：位置情報付きツイートを利用した観光ルート推薦，情報処理学会研究報告，データベース・システム研究会報告，2013-DBS-158（28），pp.1-6, 2013-11.

[8]倉田彩子，植原啓介，村井純：Twitterを用いた状況検知システムの設計と構築，情報処理学会第75回全国大会，pp.97-99, 2013-3.

[9]Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo：Earthquake shakes Twitter users: real-time event detection by social sensors, WWW 2010, pp.851-860, 2010.

[10]飯田俊彰，中村良太，後藤章：移動平均を用いた各地降水量の季節変動特性の分析，農業土木学会誌，Vol.56, No.4, pp.329-335, a1, 2011-8.

[11]安田征吾，廣川佐千男：短期・中期移動平均線を用

いた株価の解析，情報処理学会研究報告数理モデル化と問題解決（MPS），2005，37 (2005-MPS-054)，pp.23-26，2005-5.

[12]Twitter Developers：Twitter Developer 公式サイト，入手先 https://dev.twitter.com/，閲覧(2014-3).

[13]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，27巻，4号，pp.424-431，2012.

[14]独立行政法人農業・食品産業技術総合研究機構：簡易逆ジオコーディングサービス，入手先 http://www.finds.jp/wsdocs/rgeocode/index.html.ja，閲覧(2014-3).

[15]MeCab: Yet Another Part-of-Speech and Morphological Analyzer，入手先 http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html，閲覧(2012-4).

[16]気象庁：気象庁防災情報XMLフォーマット情報提供ページ，入手先 http://xml.kishou.go.jp/，2011.

[17]気象庁：さくらの観測，入手先 http://www.data.jma.go.jp/sakura/data/sakura2012.pdf，閲覧(2015-1).

[18]気象庁：かえでの観測，入手先 http://www.data.jma.go.jp/sakura/data/kaede2010.pdf，閲覧(2014-10).

[19]気象庁：いちょうの観測，入手先 http://www.data.jma.go.jp/sakura/data/ichou2010.pdf，閲覧(2014-10).

[20]株式会社ウェザーニューズ：桜情報，入手先 http://weathernews.jp/koyo/，閲覧(2015-1).

[21]公益財団法人日本観光振興協会：全国桜最前線，入手先 http://sakura.nihon-kankou.or.jp/，閲覧(2015-1).

[22]株式会社ウェザーニューズ：紅葉情報，入手先 http://weathernews.jp/koyo/，閲覧(2014-10).

[23]公益財団法人日本観光振興協会：全国紅葉最前線，入手先 http://kouyou.nihon-kankou.or.jp/，閲覧(2014-10).

（平成28年1月9日受付，平成28年4月18日採録）