

マイクロブログを用いた生物季節観測によるピーク期推定手法の検討 Examination of Best Time to See Estimation Method by Phenological Observations Using a Microblog

遠藤 雅樹^{†1 †2} 廣田 雅春^{†3} 大野 成義^{†1} 石川 博^{†2}
Masaki Endo Masaharu Hirota Sigeyoshi Ohno Hiroshi Ishikawa

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、多種多様な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、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が集まっている。SNS の 1 つであるマイクロブログは、更新が容易であるため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、実世界を表すリアルタイム性のある情報源として注目されている。マイクロブログを提供する SNS として代表的である Twitter[1]は、日本国内での利用者も多く大量の情報発信が行われている。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によると、13 歳～59 歳では 5 割を超え利用が拡大している[2]。

SNS を通して利用者が大量の情報を発信する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、SNS の活用として有用であり、実世界のリアルタイムな状況に対応した情報提供が可能となる。例えば、ある地域の観光情報を提示する際に、訪れる時期に応じておすすめ情報を変化させることでより旅行者にとって、「今」必要な情報を提供できる。実世界の「今」とは、花が咲いて見頃である状況や局所的な雨がどこで降っているかなど、情報を取得する時点での季節や天候など観光や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。生物季節観測は気象庁により行われており、気象庁が配信する気象庁防災情報 XML フォーマット形式電文の 1 つである生物季節観測報告気象報 XML により観測情報を入手することは可能である[3]。しかし、気象庁による観測は全国の気象官署の標本木を基準に行われており、同一都道府県であっても見頃には地域差があるため個々の観光スポットに対しての情報提供は困難である。また、気象情報会社や公益社団法人などが提供するサービスを利用することで観光スポットごとの見頃情報を取得することも可能である[4][5]。しかし、桜や紅葉の見頃を情報収集する場合には十分であるが、その他の生物の見頃については観測が行われていない場合や有償となるためコストがかかる。

そこで、本研究では、観光の重要な要素となる生物季節に関する情報に着目し、Twitter の位置情報付きツイートを利用して生物季節観測における見頃(ピーク期)を低コストに取得する手法を検討した。季節変化に関連した生物名を含むツイートから日本国内の生物季節変化を追跡し生物の見頃の推定を行う提案システムにより、日本国内の季節変化

をマイクロブログから取得可能となると考えられる。

本稿では、位置情報付きツイートを分析し、2014 年の紅葉シーズンの日本国内の季節変化が Twitter 内でどのように推移したのかを追跡し、実データである気象庁の公開する生物季節観測報告気象報 XML と比較した実験結果について記述することで提案手法による生物季節観測の可能性について示す。本稿は、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[6]。Kleinberg[7]は、時系列データにおいてキーワードの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。このバースト検出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[8]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[9]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[10]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のどんなイベントが盛り上がっているかを知ることができる。しかし、生物季節観測に着目しピーク期を推定する手法は検討されていない。我々は、Twitter の位置情報付きツイートから生物季節観測に関連したツイートから生物季節のピーク期を推定する手法を検討し、実世界のリアルタイムな状況に合わせた観光情報提供に結び付けたいと考えている。

3. 提案手法

本章では、日本国内の季節変化を Twitter から取得する提案手法について記述する。

我々は、Twitter から発信されたツイートの中で、日本国内の緯度経度情報を含む位置情報付きツイートを対象とした。対象とした位置情報付きツイートの本文内に季節変化に関連した生物名が記述されたツイートを抽出し分析を行った。季節変化に関連した生物名とは、気象庁において生物季節観測が行われている生物とした[11]。表 1 に、気象庁が配信する気象庁防災情報 XML フォーマット形式電文の 1 つである生物季節観測報告気象報 XML において生物季節観測対象となっている項目の一例を示す。観測対象項目は、

^{†1} 職業能力開発総合大学校 Polytechnic University

^{†2} 首都大学東京 Tokyo Metropolitan University

^{†3} 大分工業高等専門学校 National Institute of Technology, Oita College

全 68 種類であり、生物によっては複数の状態を観測する場合や複数種を同じ生物観測対象としている場合もある。例として、「かえで」は、紅葉日と落葉日の 2 項目を観測対象とし、「イロハカエデ」・「ヤマモミジ」・「オオモミジ」・「イタヤカエデ」を「かえで」として観測している。ここで、気象庁が行っている「かえで」や「いちよう」の観測条件について述べる。気象庁の公式サイトによると、「かえでの紅葉日」[12]とは、標準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。また、「いちようの黄葉日」[13]とは、標準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。「かえで」・「いちよう」のそれぞれの落葉日は、約 80%の葉が落葉した最初の日となる。

我々は、(1)対象語によるトレンド推移と(2)対象語に共起する特定語によるピーク期推定の 2 つの手法を提案する。

(1)対象語によるトレンド推移は、表 1 に示した生物名を対象語とし、対象語の漢字・ひらがな・カタカナ表記を含む位置情報付きツイートの出現頻度と時間別(日付別)推移及び位置情報から日本国内での対象語のトレンド変化を検出した。位置情報付きツイートの緯度・経度情報は、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を利用し都道府県、市区町村名を取得した[14]。

表 1 生物季節観測対象の一例

Code	Class	ClassName
01	うめの開花日	ハクバイ
		ブンゴウメ
02	つばきの開花日	ヤブツバキ
03	たんぽぽの開花日	タンポポ
		セイヨウタンポポ
04	さくらの開花日	ソメイヨシノ
		エゾヤマザクラ
		チシマザクラ
		ヒカンザクラ
05	さくらの満開日	ソメイヨシノ
		エゾヤマザクラ
		チシマザクラ
		ヒカンザクラ
<<中略>>		
13	いちようの黄葉日	イチョウ
14	いちようの落葉日	イチョウ
15	かえでの紅葉日	イロハカエデ
		ヤマモミジ
		オオモミジ
		イタヤカエデ
16	かえでの落葉日	イロハカエデ
		ヤマモミジ
		オオモミジ
		イタヤカエデ
<<中略>>		
64	にいちいぜみの初鳴日	ニイニイゼミ
		クロイワニイニ
		ミヤコニイニ
65	にほんあまがえるの初鳴日	ニホンアマガエル
66	にほんあまがえるの初見日	ニホンアマガエル
67	はるぜみの初鳴日	ハルゼミ
68	みんみんぜみの初鳴日	ミンミンゼミ

表 2 対象語一覧

項目	対象語
かえで	かえで, 楓, カエデ
いちよう	いちよう, 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

また、比較対象とした気象庁の配信する気象庁防災情報 XML フォーマット形式電文の 1 つである生物季節観測報告気象報 XML を受信しデータベースに蓄積するシステム運用を 2014 年 7 月末から開始した。2014 年 1 月から 7 月までの不足分は、気象庁から受信したデータを蓄積し公開を行っているサイトを利用し補完した[15]。実験により取得した位置情報付きツイートと比較対象とした気象庁の生物季節観測データは、ウイングアーク 1st 株式会社の MotionBoard[16]の地図機能を利用した可視化を行い位置情報付きツイートによる生物季節観測の可能性を確認した。

次に、(2)ピーク期推定手法について記述する。ピーク期推定は、表 2 に示す対象語と対象語に共起する特定語を利用した。対象語に共起する特定語とは、「きれい」・「綺麗」・「キレイ」の 3 語と「東京」や「京都」などの地名とした。この特定語は、対象語との相関を求め正の強い相関が得られた語から今回は人手で選択した。自動的に共起語を選ぶ手法についてはさらなる議論が必要であると考えている。また、一般にツイートなどテキストデータを利用する際には、ノイズ除去などの処理が必要となる。ツイートにノイズが混じることで推定精度を損なう原因になるが、我々は、ツイート観測期間の全体平均を閾値として全体平均以下をノイズと判断することとした。この手法によりツイートのほとんどがノイズである時期をピーク期前後の定常状態、全体平均を超える時期をピーク期と判断することとした。そして、期間推定には、特定語の日ごとの推移から 7 日移動平均と 30 日移動平均を求め、7 日移動平均が 30 日移動平均を 5 日間連続で上回った日をピーク開始、7 日移動平均が 30 日移動平均を下回った日をピーク終了とした。ここで、移動平均の間隔は、位置情報付きツイート数は土日に増加し週単位で推移する傾向を予備実験により取得したことから、週単位の 7 日と週単位に対して長期間となる月単位の 30 日を設定した。移動平均による特定語の推移から対象語とした生物の見頃のピーク期推定の可能性を確認した。さらに、気象庁の観測データの地図上での推移と対象語に共起する地名を含むツイートの推移から地域ごとのピーク期推定の可能性についても検討した。

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行った検証実験の実験方法と実験結果を示す。4.1 節に(1)対象語によるトレンド推移の実験方法、4.2 節に実験結果を示す。また、4.3 節に(2)ピーク期推定の実験方法、4.4 節に実験結果を示す。

4.1 対象語によるトレンド推移の実験方法

本実験で使用したデータセットは、2014 年 6 月 11 日から 12 月 19 日までの日本国内における位置情報付きツイートを利用した。対象期間のデータ数は、約 7,000 万件である。次に、表 2 に本実験での対象語一覧を示す。対象語は、9 月から 12 月に気象庁の観測官署において観測される秋の植物である「かえで」と「いちよう」に加え、一般に秋の観光対象として用いられる表現である「こうよう」の 3 種類とした。比較対象とした気象庁の生物季節観測対象の項目は、「かえでの紅葉日」・「かえでの落葉日」・「いちようの黄葉日」・「いちようの落葉日」の 4 項目とした。本実験において、気象庁の観測データのトレンド推移と位置情報付きツイートに対象語が含まれるツイートのトレンド推移を確認し、位置情報付きツイートによる生物季節観

測の可能性を検証した。

4.2 対象語によるトレンド推移の実験結果

表3に気象庁の観測する「かえで」の紅葉日・落葉日と「いちよう」の黄葉日と落葉日のトレンド推移を示す。各観測地点における「かえでの紅葉日」と「かえでの落葉日」の差は平均で約5.2日となった。また、各観測地点における「いちようの黄葉日」と「いちようの落葉日」の差は平均で約6.1日となった。それぞれの平均は、紅葉日・黄葉日と落葉日の両方の観測が行われた地点のみで求めた。

次に、表3に示した気象庁の観測データについて、図1に「かえでの紅葉日」と「かえでの落葉日」の日付での推移、図2に「いちようの黄葉日」と「いちようの落葉日」の日付での推移を示す。

図1より「かえで」は、10/10に紅葉が観測(北海道釧路市)され、11/26に紅葉の観測ピークがある。落葉は、10/19に観測(北海道釧路市)され、12/1に落葉の観測ピークがある。図2より「いちよう」は、10/23に黄葉が観測(北海道旭川市)され、11/20に黄葉の観測ピークがある。落葉は、11/3に観測(北海道函館市)され、12/1に落葉の観測ピークがある。この結果から、本実験の比較対象とした気象庁の観測データから2014年の「かえで」のピーク期を11/26-12/1、「いちよう」のピーク期を11/20-12/1と想定して検証を行うこととした。

次に、図3に位置情報付きツイートに含まれた対象語の推移を示す。図3の第1軸は「かえで」・「いちよう」、第2軸は「こうよう」のツイート数を表す。期間中に観測したツイート数は、「かえで」・「いちよう」・「こうよう」それぞれで、7,993件・4,708件・20,017件である。

「かえで」は、11/30付近にピークを観測し、気象庁の「かえでの紅葉日」と「かえでの落葉日」の観測ピーク期間であった11/26-12/1の期間中にピークを確認できたが、実験期間全体において明確なトレンド変化を確認することができなかった。これは、「かえで」「楓」を含む人名などを表現するツイートの割合が自然現象の「こうよう」を表す表現より多く利用されており、気象庁の観測する「紅葉」期間であっても人名を表すツイート量が多くツイート数だけではトレンド変化を捉えることができないことを指していると考えられる。「いちよう」の対象語ツイートは、「かえで」同様に11/30付近にピークを確認した。気象庁の「いちようの黄葉日」と「いちようの落葉日」の観測ピーク期間である11/20-12/1の期間中にピークを確認でき、実験期間内でも11/30前後にトレンド変化を確認することができた。また、「こうよう」の対象語ツイートは、11/23付近にピークを観測した。気象庁の「かえで」と「いちよう」の観測ピーク期間である11/20-12/1の期間中にトレンド変化を確認することができ、実験期間内でも11/23前後のトレンド変化を確認できた。この結果から、位置情報付きツイートの対象語ツイート件数によるトレンド変化は、自然現象の「紅葉」「黄葉」に関連した用途に用いられることが多いキーワードである「いちよう」や「こうよう」では観測可能であると考えられる。しかし、自然現象の「紅葉」「黄葉」を表す用途以外にも使われる機会が多い「かえで」についてはトレンド変化をツイート数でのみ観測することは困難なため、以降に記述する対象語と共起する特定語などを利用した分析が必要であると考えられる。

表3 気象庁観測データ

観測地	かえで 紅葉日	かえで 落葉日	いちよう 黄葉日	いちよう 落葉日	観測地	かえで 紅葉日	かえで 落葉日	いちよう 黄葉日	いちよう 落葉日
北海道釧路市幸町	10/10	10/19	-	-	茨城県水戸市金町	-	-	11/20	-
北海道帯広市東4条南	10/18	10/27	10/23	11/3	大分県大分市長浜町	11/21	12/1	12/1	-
北海道旭川市宮前通東	10/19	10/23	-	-	三重県津市島崎町	11/21	12/3	11/21	-
北海道函館市美原	11/2	11/11	10/28	11/3	京都府京都市中京区	12/9	-	11/21	11/29
北海道札幌市中央区北2条西	10/30	11/3	10/30	11/3	高知県高知市本町	-	-	11/21	12/2
岩手県盛岡市山王町	11/16	11/25	10/30	11/10	和歌山県和歌山市男野芝丁	12/11	-	11/21	11/30
青森県青森市花園	11/6	11/20	10/30	11/7	香川県高松市伏石町	11/24	12/1	11/22	11/27
長野県長野市箱清水	11/19	11/29	11/2	11/17	山梨県甲府市飯田	12/1	12/3	11/22	12/3
北海道室蘭市山手町	11/3	11/4	-	-	群馬県前橋市昭和町	-	-	11/23	-
秋田県秋田市山王	11/7	11/14	11/3	11/6	富山県富山市石坂	11/27	12/1	11/23	11/26
兵庫県神戸市中央区	11/25	12/7	11/4	11/25	佐賀県佐賀市駅前中央	11/25	11/29	11/23	11/29
福島県福島市松木町	11/21	-	11/7	11/18	宮崎県宮崎市霧島	12/3	-	11/24	12/1
広島県広島市中区上八丁堀	11/18	12/2	11/10	11/25	鹿児島県鹿児島市東郡元町	-	12/5	11/24	12/1
山形県山形市緑町	11/12	12/1	11/12	11/21	東京都千代田区大手町	11/25	-	11/25	-
石川県金沢市西念	11/26	11/29	11/13	11/21	奈良県奈良市半田開町	11/26	12/2	11/25	12/1
愛知県名古屋市中区日和田	11/26	-	11/13	11/28	栃木県宇都宮市明保野町	11/26	12/6	12/3	12/8
新潟県新潟市中央区美咲町	11/16	11/30	-	11/26	岐阜県岐阜市加納二之丸	11/26	12/8	11/27	-
滋賀県彦根市城町	12/1	12/11	11/17	-	徳島県徳島市大和町	11/26	12/1	-	-
島根県松江市西津田	11/21	12/1	11/18	11/25	鳥取県鳥取市吉方	11/26	12/3	11/28	12/8
福井県福井市豊島	11/25	12/2	11/19	11/25	熊本県熊本市西区春日	11/27	12/3	-	-
愛媛県松山市北持田町	12/11	-	11/19	11/29	福岡県福岡市中央区大濠	-	-	11/28	12/11
静岡県静岡市駿河区曲金	11/28	-	11/20	-	埼玉県熊谷市桜町	11/29	-	12/2	12/10
岡山県岡山市北区桑田町	11/27	12/4	11/20	12/4	神奈川県横浜市中区山手町	-	-	12/2	-
宮城県仙台市宮城野区五輪	11/22	11/29	11/20	12/2	千葉県銚子市川口町	-	-	-	12/7
大阪府大阪市中央区大手前	11/27	12/5	11/20	11/30	東京都千代田区北の丸公園	-	-	-	12/8
山口県下関市竹崎町	12/3	-	11/20	12/1	長崎県長崎市南山手町	12/10	-	12/10	-

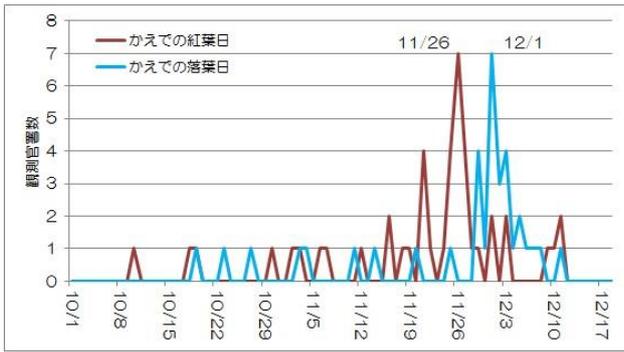


図1「かえで」の観測官署数推移



図2「いちょう」の観測官署数推移

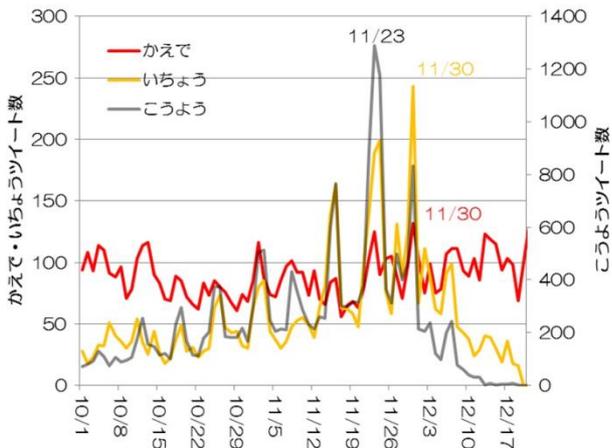


図3対象語数推移

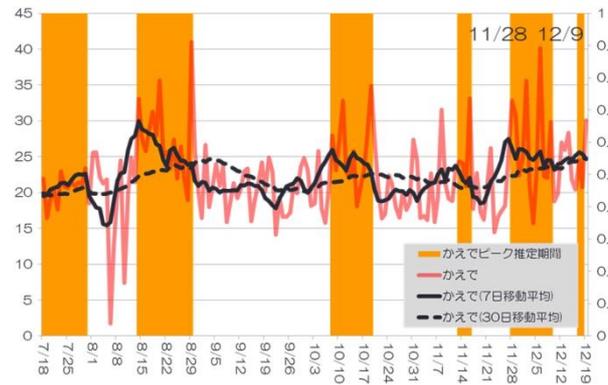


図4「かえで」ピーク期推定結果



図5「いちょう」ピーク期推定結果

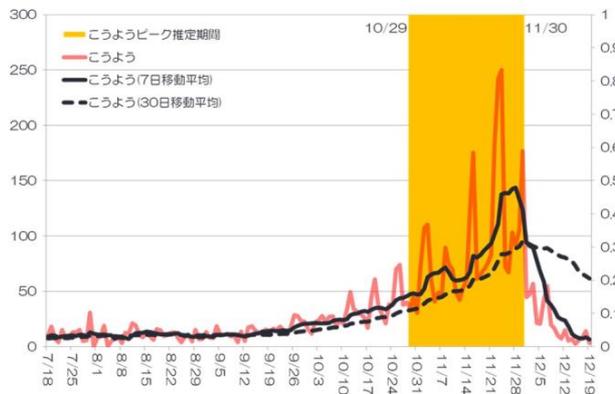


図6「こうよう」ピーク期推定結果

4.3 ピーク期推定の実験方法

4.2節に示した位置情報付きツイートに対象語を含むツイートに対して移動平均法を用いた分析によるピーク期推定の実験を行った。移動平均法を用いたピーク期推定の実験対象は、(1)表2の対象語、(2)3節に示した「きれい」・「綺麗」・「キレイ」の3つの特定語とした。また、対象語に共起する地名の推移と気象庁の観測データの比較実験を行い地域ごとのピーク期推定の可能性についても実験を行った。

4.4 ピーク期推定の実験結果

対象語である「かえで」・「いちょう」・「こうよう」を用いたピーク期推定結果をそれぞれ図4・図5・図6に示す。各図の黒実線は7日移動平均、黒点線は30日移動平均

均、色付き実線はツイート数の推移を表し、橙色部分が提案手法により推定したピーク期である。図4の「かえで」は、11月に注目すると11/28-12/9をピーク期と推定し、気象庁の観測データから想定したピーク期の11/26-12/1に近似したピーク期推定ができた。しかし、11月下旬以外にも複数のピーク期が推定され、「かえで」についてはピーク日の観測と同様に対象語のみでの推定は困難であるとの結果になった。図5の「いちょう」はピーク期を11/11-12/4と推定し、気象庁のデータからの想定は11/20-12/1に近似した推定結果となった。図6の「こうよう」はピーク期を10/29-11/30と推定した。「こうよう」は、「かえで」・「いちょう」のどちらの表現でも使われる語のため推定期間が長くなったと考えられる。

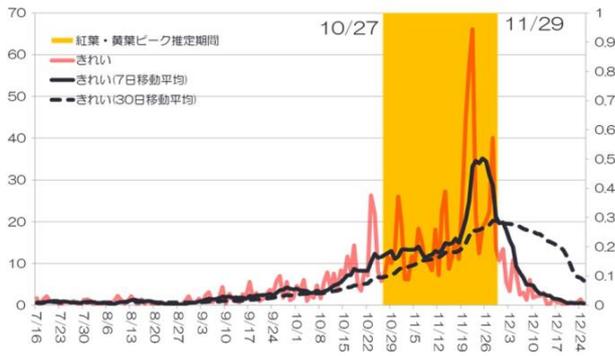


図 7 特定語によるピーク期推定結果



図 8 紅葉・黄葉観測分布(2014/11/2-11/8)

次に、図 7 に表 2 に示す対象語「紅葉」に共起し、「紅葉」の推移に強い正の相関が得られた「きれい」・「綺麗」・「キレイ」の 3 つの特定語を含むツイートを利用したピーク期推定結果を示す。特定語によるピーク期推定は 10/27-11/29 となった。対象語に共起する語から正の強い相関が得られた語を対象とすることで、「こうよう」に近いピーク期推定を行うことができた。「黄葉」に共起し強い正の相関が得られた「綺麗」・「きれい」・「黄色」を利用したピーク期推定も図 7 とほぼ同様の結果を示した。この結果から、対象語に共起する特定語がほぼ同じであったことからピーク期推定に差が出なかった。しかし、「かえで」など対象語のみでは紅葉ツイートとしてピーク期推定ができない場合であっても「紅葉」と一般的に「かえで」の表現として利用される語の共起語を利用することでピーク期推定が行える可能性を確認した。

次に、気象庁観測データの地図上でのトレンド変化と対象語に共起する地名を含むツイートの推移との比較について記述する。4.2 節の図 1 と図 2 で示した「かえで」・「いちよう」に関する気象庁の観測データについての地図上での推移を図 8 から図 10 に示した。図 8 から図 10 はそれぞれ、11/2-11/8、11/16-11/22、11/23-11/29 の期間での状況である。各図のポイントは気象庁において観測された観測官署を表し、赤色：「かえでの紅葉日」、黄色：「いちようの黄葉日」、橙色：「かえでの紅葉日」かつ「いちようの黄葉日」である状態を示す。各ポイントで落葉日を観測した時点でポイントを除去した。表 3 に示すとおり

10/10 に釧路地方気象台(北海道釧路市幸町)で「かえでの紅葉日」が観測されてから徐々に南下していく様子を確認できる。また、図 11 に対象語に共起する地名を含むツイートの推移を示す。図 11 に示す対象語と共起する地名でのツイート数のピークは、北海道・会津・大阪・京都・東京都・鎌倉のそれぞれで、10/10・11/7・11/21・11/22・11/28・11/28 となった。気象庁による各地域の観測日は、「かえで」・「いちよう」で観測日が早い日を基準とすると、北海道・福島県・大阪府・京都府・東京都・神奈川県それぞれで、10/10・11/7・11/20・11/21・11/25・12/2 であった。この結果から、位置情報付きツイートに含まれる対象語と共起する地名の推移は気象庁の観測日とほぼ一致した。このため、対象語に共起する地名ごとのツイート数から各地域のピーク期を推定可能と考えられる。しかし、図 11 に示したとおり、地名によってツイート量が大きく異なるが地域ごとにピークが異なることが確認できる。そこで、共起した地名別のツイートを利用し提案手法によるピーク期推定実験を行った。最も共起語として出現回数が少なかった「会津」を図 12、最も出現回数が多かった「東京都」を図 13 に示す。限定したピーク期推定のため、地域によってはデータ量が少なくピーク期推定が困難であることを確認した。



図 9 紅葉・黄葉観測分布(2014/11/16-11/22)



図 10 紅葉・黄葉観測分布(2014/11/23-11/29)

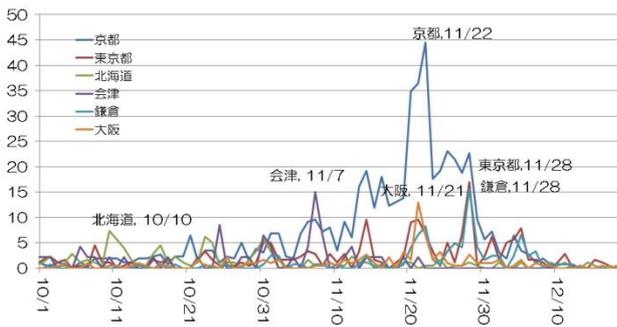


図11 対象語に共起する地名を含むツイートの推移

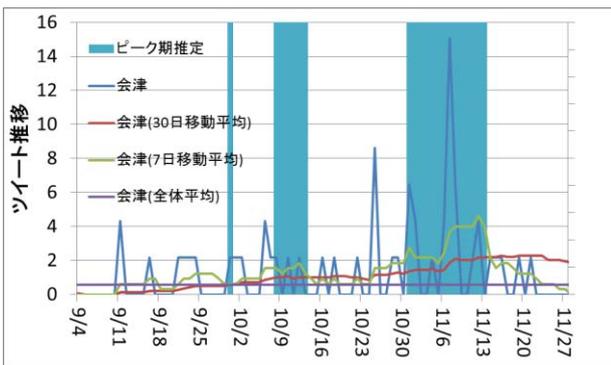


図12 対象語に共起する地名を含むツイートの推移

5. おわりに

本稿では、位置情報付きツイートに含まれる「かえで」・「いちょう」・「こうよう」に関する語を対象語として、日ごとの時間的変化を分析し、気象庁の生物季節観測データと比較を行った。その結果、紅葉・黄葉に関する気象庁の観測日と同等のツイート数のピークを観測可能であることを確認し、紅葉・黄葉に関するツイートの盛り上がりを実世界での季節変化には関連があり、生物に関連するツイートの観測を行うことで見頃(ピーク期)を推定できる可能性があることを確認した。さらに、対象語とした「かえで」

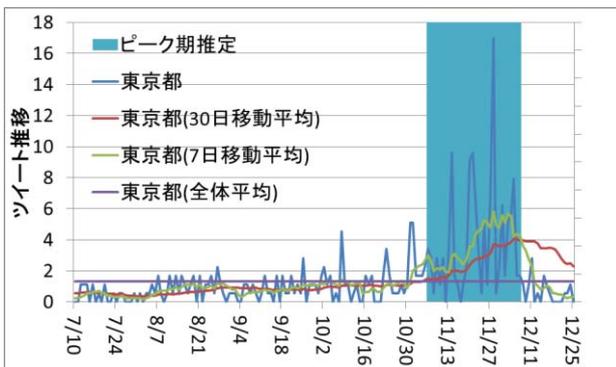


図13 対象語に共起する地名を含むツイートの推移

・「いちょう」・「こうよう」などの生物名に共起する「綺麗」・「鮮やか」などの特定語を利用し、移動平均法によるピーク期推定手法を提案した。提案手法を用いた実験により、紅葉・黄葉に関連したツイートからピーク期推定が可能であることを確認した。よって、本稿で提案した手法を用いた生物季節観測の見頃(ピーク期)を推定することは可能であると考えられる。

しかし、本稿で実験対象とした位置情報付きツイートのデータ量は、都市圏に集中し郊外には少ない傾向がある。このため、本稿では、地域ごとに十分な位置情報付きツイート量が集まらなかったことから、提案手法を用いて位置情報付きツイートを分析することで詳細な地域ごとのピーク期推定は可能であることを確認したに留まった。今後は、位置情報付きツイート以外に位置情報無しツイートや Flickr などデータ量を確保する手法を検討し、より精度の高いピーク期推定につなげられるよう研究を進めていく予定である。

参考文献

- [1] Twitter, <https://twitter.com/>.
- [2] 総務省, "平成25年通信利用動向調査の結果", http://www.soumu.go.jp/johotsusintokei/statistics/data/40627_1.pdf, 2014.
- [3] 気象庁, "気象庁防災情報 XML フォーマット情報提供ページ", <http://xml.kishou.go.jp/>, 2011.
- [4] 株式会社ウェザーニューズ, "紅葉情報", <http://weathernews.jp/koyo/>.
- [5] 公益財団法人日本観光振興協会, "全国紅葉最前線", <http://kouyou.nihon-kankou.or.jp/>.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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), 1-6, 2013-11.
- [10] 倉田彩子, 植原啓介, 村井, "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年全国大会, pp.97-99, 2013-3.
- [11] 生物季節観測対象, <http://www.data.jma.go.jp/sakura/data/index.html>.
- [12] かえでの観測, <http://www.data.jma.go.jp/sakura/data/kaede2010.pdf>.
- [13] いちょう観測, <http://www.data.jma.go.jp/sakura/data/ichou2010.pdf>.
- [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, <http://www.nds.jp/wsdocs/rgeocode/index.html>.
- [15] mk-mode, <http://www.mk-mode.com/rails/jmaxml/db/seibutsu>.
- [16] ウイングアーク 1st 株式会社 MotionBoard, <http://www.wingarc.com/product/motionboard/>.