

Web からの地域・観光情報収集とその有用性の検討

松本義之、藪内賢之

1. はじめに

近年、地域情報あるいは観光・行楽情報を収集する手段として、個人が Web ページから収集する傾向が強くなってきている。これまでは、旅行代理店を来訪して店員から旅行情報を聞く場合や、雑誌やテレビなどから情報を収集することが多かった。しかし、コンピュータ・携帯電話・スマートフォンなどが普及するようになり、自分で Web を利用した情報収集が可能になった。2013 年 8 月に株式会社日本リサーチセンターが行ったメディアについての調査¹⁾によると、旅行情報の入手先メディアとして、テレビが 32.8%、雑誌が 25.3%、Web が 21.3%となっている（図 1）。また、20 代・30 代においては、Web での入手が 34.6%、32.5%となり、入手先メディアの 1 位となっている（図 2）。

また、観光地を訪れた来訪者が、その体験・感想を blog や SNS などを使って情報発信している。情報技術の発展により、これまでに比べ簡単に情報を発信することが可能になっており、個人の経験を記述した「口コミ」情報が利用者・体験者の「生の声」として重視されるようになった。そのため、観光推進・地域振興を行うためには Web 上での情報発信・情報収集が不可欠である。

その一方で、Web には膨大な情報が溢れ、必要な情報を見つけ出す技術・方法が求められている。すなわち、有益な情報がシステムティックに蓄積されたのではなく、個々に情報を公開しているために膨大な情報に埋もれ、見付けることが困難な状況にある。そこで、Web ページから自動的にデータを収集・蓄積、知識・知見を発見・抽出する試みが行われている。これら技術を用いた、Web ページを構成するテキスト情報から特定パターンの発見、Web ページへのアクセスパターンの発見、Web リンク構造パターンの発見などの研究が盛んに行われている^{2) 3)}。

本研究では、Web 上に存在する膨大なデータから、観光推進や地域振興などで活用するために「下関」「関門地域」などの地域情報を収集し、当地域の特徴抽出を目的としている。インターネット上の様々な媒体から発信される情報を収集・蓄積し、分析を行う事で、様々な知見を得ることができると考える。インターネット上での情報発信手段として従来型 Web ページ・blog・SNS などがある。その中でも SNS は、個人での情報発信が容易であり、また速報性も高い。双方向性も他メディアと比較して優位である。本研究では、このような SNS の中でも、特に速報性が高く、手軽に情報発信が可能な Twitter に着目した。Twitter から発信される情報を収集・蓄積・分析することで、様々な知見を得ることを目的としている。

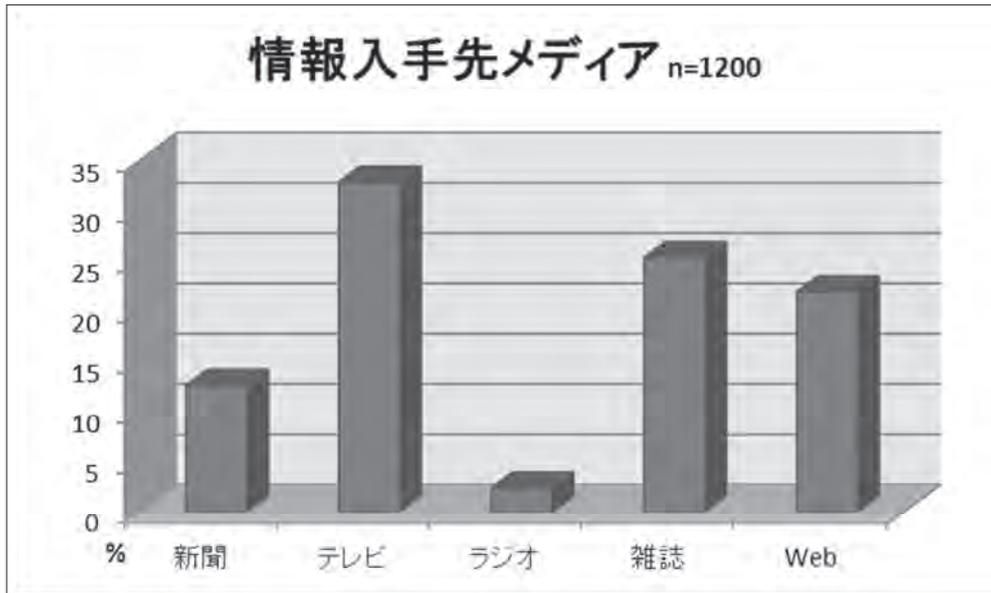


図1 旅行情報の入手先メディア（日本リサーチセンター HP¹⁾より引用）

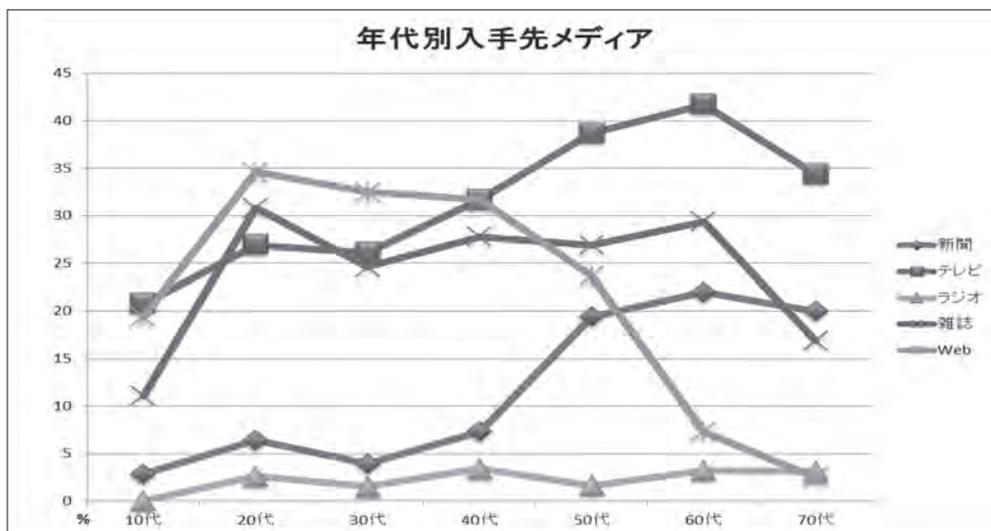


図2 年代別旅行情報入手先メディア（日本リサーチセンター HP¹⁾より引用）

2. SNS の概要

2.1 Web 情報の種類

インターネット上では、様々な形式で情報発信が行われている。その中でも代表的な媒体が、従来型 Web ページ・blog・SNS である。

従来型 Web ページとは、企業や団体の広告・宣伝用に利用されているホームページであり、インターネット初期から利用されている。従来型 Web ページはプル型情報サービスであり、ユーザが自分からホームページを訪ねていき、必要な情報を引き出す必要がある。以前は、ユーザが URL を直接入力してホームページを来訪していたが、現在は

googleなどの検索エンジンで検索を行い、目的のホームページから情報を収集している。

blogとは、主に個人で運営されている日記風のホームページである。基本的に時系列順に記事（日記）を表示する形式をとっており、blog運営者の見解や独自の情報を日記風に記載するが多い。一般的なblogは、他のblogに対してリンクを張った場合、リンク先に対してその事を通知するトラックバック機能や、来訪者の意見を投稿できるコメント機能などの双方向性を備えている。

SNSとは、人間同士のコミュニケーションを促進するコミュニティ型のホームページである。友人同士のコミュニケーションを図る事に加え、インターネット上で知り合った共通の趣味などをもつもの同士の、新しい人間関係を構築することが可能である。また、SNSはプッシュ型情報サービスであり、一度設定しておく、自動的に設定先からメッセージが送信されてくる。それ以外にもソーシャルボタン（いいね！ボタン）などの情報送信手段があり、非常に情報拡散性が高く、双方向性も高いWeb情報となっている。

株式会社ITC総研による2013年のSNS利用動向に関する調査⁴⁾によると、2012年度末日本における国内ネットユーザー9,556万人のうち、SNS利用者は4,965万人であり、ネットユーザー全体のうちの52%にあたる（図3）。また、同社が2013年5月に実施したSNS・通話アプリに関するアンケート調査では、Facebook利用者が34%、Twitter利用者が26%、mixi利用者が22%となっており（図4）日本国内においては、上記3つのSNSが非常によく利用されていることが判る。

本研究では、Web情報のなかでSNSの1つであるTwitterからの情報収集・分析を行う。観光や行楽情報をインターネットから得ようとする場合、ユーザは速報性や口コミ情報を重視していると考えられる。まとまった情報・速報性が高くない情報が必要な場合は、従来メディアである旅行雑誌や旅行代理店からの情報収集の方が適していると考えられる。Twitterは、SNSの中でも速報性が高く、手軽に口コミ情報を発信・発見できる利点がある。リアルタイムな口コミ情報を得ようとするれば、Twitterが最も適したメディアであると考えられる。

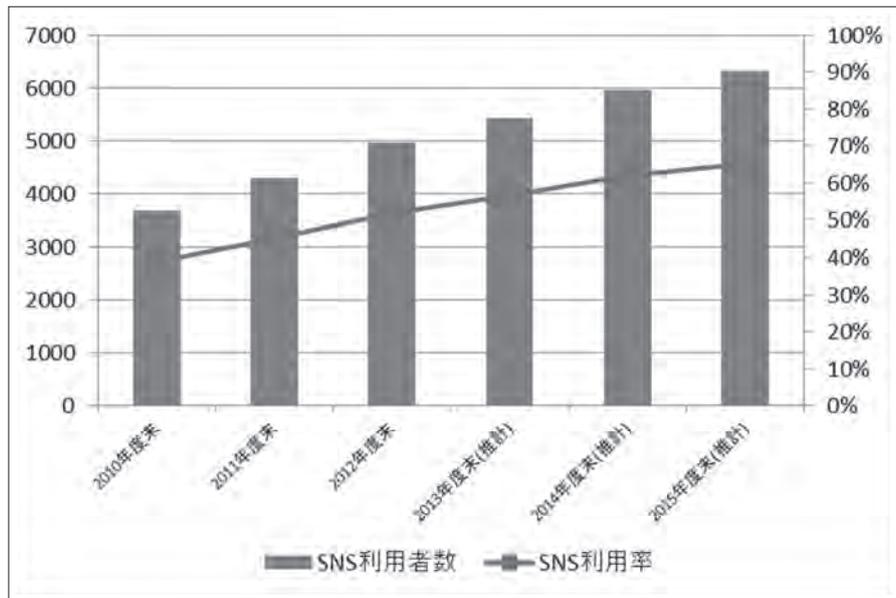


図3 SNS利用動向に関する調査 (ICT 総研 HP⁴⁾より引用)

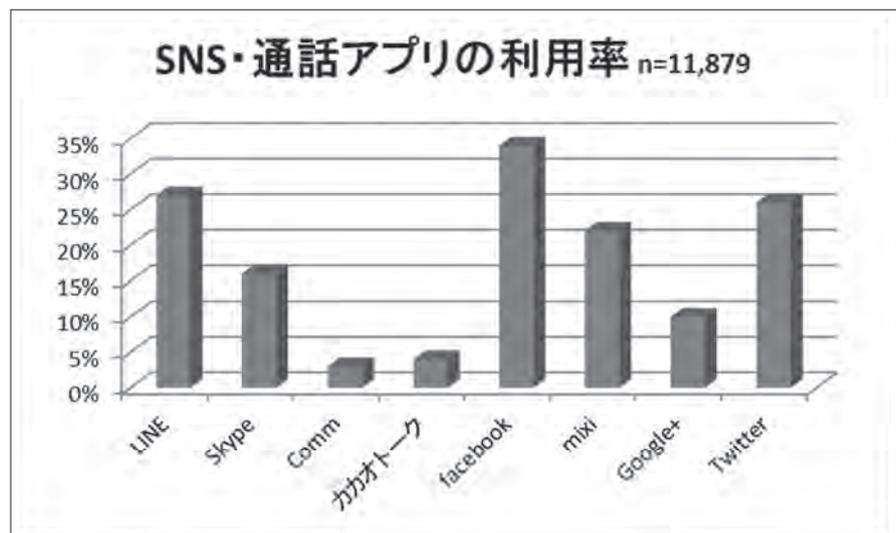


図4 SNS・通話アプリの利用率 (ICT 総研 HP⁴⁾より引用)

2. 2 Twitter について

現在、日本で認知度が高い SNS としては、Facebook、Twitter、mixi が挙げられる。その中でも Twitter は非常に情報拡散性・速報性が高いメディアと言われている特徴がある SNS である。2006 年 7 月からサービスが開始されており、Facebook などと違い、メールアドレスが異なれば 1 人で複数アカウントを作成することも可能である。また Twitter 自身の発表によると⁵⁾アクティブユーザー数は 2 億 3 千万人、1 日のツイート数は平均 5 億ツイート数に上るとの事である。

Twitter が持つ機能としては、140 字以内のメッセージ投稿（ツイート）機能、時系列

順にツイートを表示するタイムライン表示機能、他人のツイートをタイムラインに自動表示するフォロー機能、気に入ったツイートを、そのままフォロワーに転送するリツイート機能などがある。

また、Twitter を利用した企業によるマーケティング活動も数多く行われている。従来のマーケティング活動といえば、テレビ・雑誌などを利用したものが大部分であったが、インターネットの普及により、SNS を利用したマーケティング活動が行われるようになった。特に Twitter は情報発信が手軽であり、また、リツイート機能を利用した情報の拡散スピードが速いという特徴がある。このような Twitter の特徴を利用した広告方法が、「プロモアカウント」「プロモツイート」と呼ばれるものである⁶⁾。例えば、アメリカのトップミュージックレーベルである Interscope Records は、@Interscope というアカウント名（プロモアカウント）で、プロモツイートをを行っている。同レーベルに所属する Lady Gaga の新アルバム発売においては、Lady Gaga 自身の Twitter アカウントである @ladygaga からのツイートと自社のプロモツイートを組み合わせて宣伝を行い、30%という高いエンゲージ率（プロモツイートに対して、ユーザがリツイートを行った割合）を記録した。

3. WebAPI

3. 1 WebAPI の概要

WebAPI とは、インターネットを経由して利用できる API の事である。API とは、Application Program Interface の略で、通常は Windows などが提供している Win32API などの事を指す。OS が提供している API は、応用プログラムから呼び出すことが可能な機能群・関数群であり、API を利用することで、応用プログラムは、異なるハードウェア上で同一プログラムを実行させることが可能となっている。WebAPI は、OS などで提供されている API と同様の仕組みをインターネット経由で実現したものであり、Web サイトで提供されている各種機能を、別の Web サイトやクライアントアプリケーションで利用可能にしたものである。

代表的な WebAPI としては、Google 社が提供している Google Maps API⁷⁾、Amazon 社が提供している Amazon Product Advertising API (Amazon Web サービス)⁸⁾がある。Google Maps API は、Google 社が提供している地図情報 API であり、オリジナルの位置情報アプリケーション作成や、自分のホームページに地図情報を追加することが可能になる。Amazon Product Advertising API は、Amazon 社が提供する Amazon の商品情報を表示するための API であり、取扱商品の検索や人気商品一覧などを取得可能である。

3. 2 TwitterAPI

本研究で利用しているのは、Twitter 社が提供している WebAPI である TwitterAPI である。TwitterAPI は、クライアントアプリケーションや Web ページからツイートの投稿や検索などを可能にする。この TwitterAPI は、2013 年 6 月 12 日より、バージョン 1.0 か

らバージョン 1.1 に変更された⁹⁾。この変更は、悪意のあるアプリケーションや非常に高いレートで API を利用する bot（自動的にツイートをを行うアプリケーションプログラムの総称）などの排除を目的としている。そのため、これまで不要であったリクエストに関する認証手続きが必要となり、API コール限度数も引き下げられている。認証手続きには OAuth 認証が使用されている。

OAuth 認証とは、ユーザ ID・パスワードを受け渡しせずに、サービスとアプリケーション間でデータの受け渡しを可能とする認証方式である。通常のパスワード認証方式と異なり、アプリケーションにユーザ ID・パスワードを開示する必要がないため、悪意のあるアプリケーションがユーザ ID・パスワードを収集し、不正利用を行うことを防ぐことができる。

Twitter において OAuth 認証を行うためには、事前に開発用アプリケーションを登録しておく必要がある。登録のためには、アプリケーションの URL・許可される動作などの設定を行う。登録を行うことで、Twitter から、Access token・Access token secret を取得することができる。取得した Access token・Access token secret を用いて、アプリケーションから OAuth 認証を行うと、TwitterAPI を利用することが可能になる。Twitter から検索結果を取得するためには、GET search/tweets を利用する。このようにして Twitter に対して、アプリケーションから任意の検索結果を引き出すことができるようになる。

4. Web マイニング

4. 1 Web マイニングとは

Web マイニングとは、データマイニングの一種である。データマイニングとは、データ分析手法の 1 つであり、データベース・データウェアハウス（分析専門のデータベース）の中から有益な情報を引き出すことを目的としている。有名な事例として、「おむつとビール」の併せ買いがある。一般に「アメリカの大手スーパーマーケットチェーンにおいて、POS データを分析した結果、若い男性が週末の夕方におむつとビールを一緒に買う傾向にある」という内容で知られている。この事例は、1992 年 12 月 23 日の The Wall Street Journal に掲載された "Supercomputer Manage Holiday Stock" という記事が発端と言われているが、実際にあった事例かどうかは疑問視されている。しかしながらデータマイニングの意義と可能性について、よく説明された事例である。データマイニングは、上記例のようにデータベース・データウェアハウス上に整えられたデータに対する分析手法であるが、Web マイニングは、インターネット上にあるデータを分析対象とする。データベースに保存されているデータは、形式的に整えられたデータであり、比較的分析が容易であるが、Web 上のデータは、大部分がテキストデータであり、形式的にも整えられていない。こういったデータを分析するためには、テキストマイニングと呼ばれる手法を用いる必要がある。

4. 2 テキストマイニング

インターネット上に保存されているデータの大部分はテキストデータであり、これらのデータを分析しようとするれば、テキストマイニングと呼ばれる手法を用いることになる。テキストマイニングとは、小説・新聞・Web ページなどの文章を対象とするデータ分析技術である。テキストマイニングを用いる目的は、大量に存在するテキストデータの中から、特定のテーマを抽出することや、文章の分類を行うこと¹⁰⁾であり、様々な研究が行われている。テキストデータの総量が少なければ、コンピュータを用いずに分析を行うことも可能であるが、インターネット上に保存されているような大量データを分析する場合は、必要不可欠な技術であるといえる。実際に様々なデータ分析が行われており、インターネット上における blog 情報を分析し、即時性・重要性が高い情報を発信する有用な blogger を発見するための分析¹¹⁾や、インターネット上でユーザが感心を持っている Web サイトや検索キーワードを抽出して、視覚化を行う研究¹²⁾などが行われている。

5. 地域・観光情報の収集・分析

5. 1 地域・観光情報

欧州原子核研究機構（CERN）のティム・バーナーズ・リーらが 1989 年 3 月に World Wide Web の格子を築いて 25 年が経った。これまで紙媒体に印刷されてきた情報はデジタル化され、ハイパーテキストという形態となってインターネットで公開されるようになった。さらに、情報伝達の手段・方法が発展したため、多くの分野で情報のデジタル化が進んだ。

これまで、インターネットで情報公開するには特別な設備を保有しなければならなかったが、特別な設備や費用を費やすこともなく、簡単に各種情報を公開することが可能となった。以前は、主に企業や各種団体が Web ページを公開していたが、個人ユーザが Web ページを公開している。そして、個人的な覚え書きや日記などの blog、個人同士の繋がりを促進・サポートする SNS（Social Networking Service）へと発展した。これにより、これまで入手することが困難であった口コミ情報へのアクセスが容易になった。

誰でも容易にインターネット上で情報公開できるようになったため、SNS への書き込み件数も爆発的増加の傾向がある。その情報は、主に日常の出来事である。その書き込み情報は、特定の人達にとっては宝である。例えば、地域振興や観光振興に携わる者にとっては、blog や SNS などに書き込まれた情報は注目されている事柄であり、イベントや観光資源を企画し、作り出す手掛かりとなる。

このため、近年では色々な分野で SNS に書き込まれた情報を活用することが試みられている。このようなビッグデータからの情報収集および知識発見は容易ではなく、課題も多い。それは膨大なデータから有益な情報を発見するための処理時間と労力を要すること、膨大なデータを効率的に処理する技術などである。

5. 2 地域・観光情報の収集

これまで、地域情報や観光情報にアクセスすることは容易ではなかったが、近年はインターネット・サービスが発達し、blog や SNS などに各種口コミ情報が書き込まれるようになった。このため、本研究では Web から地域・観光情報の収集を目的とし、Twitter を情報源にする。blog は日記のような形態となっているため情報量は豊富であるが、該当する Web ページを見つけ出す時間、Web マイニングのための処理時間などを考慮するとデータ発見の対象とするのは現実的ではない。Twitter は一件 140 文字以内のメッセージで、SITE 内の検索に制限はあるものの blog と比べればデータ収集は効率的である。このため、本研究では Twitter の書き込みから地域・観光情報を収集することにする。

分析には次の2つのデータを対象とした。

- 長期データ：2013年7月9日～2014年4月15日（5,646件、リツイート含まず）
- 短期データ：2014年3月5日14時～3月6日14時（1,478件、リツイート含む）

これらは下関に関するツイートで、長期データは5,646件あり、リツイートを含まない。また、短期データは1,478件で、リツイートを含める。長期データでリツイートを含まなかったのは、数が大きくなりすぎて処理できないからである。

ここで、長期データでインターネット・ユーザの関心事を見付け、短期データで関心の大きさを確認することを目指している。このため、本分析では次の分析手順を取る。

長期データ

目的：長期間の調査で、ユーザの関心事を調べる。

Step 1：3つのタームで共起を確認する。

ツイッターに書き込まれている3つのタームを見付ける。

Step 2：繋がりを確認する。

- ・ Step 1で見付けたタームを基準にして、それと共に用いられるタームを調査する。
- ・ また、それと共に用いられるタームを調査する。

短期データ

目的：長期データで見付けた関心事を確認する。

Step 1 と Step 2 は同じ。

5. 3 ツイートの解析

テキストマイニングの場合、文意に関係ない記号が含まれている。このため、一般的にはデータをクリーニングしてから形態素解析・構文解析を行い、分析する。しかし、ツイートの場合、ハッシュタグや顔文字などが文章に多数含まれており、最初にツイートをクリーニングするのでは効率が悪い。このため、形態素解析・構文解析の後にクリーニングし、

分析することにした。

言語で意味をもつ最小の文字列の単位を形態素といい、文を単語に分割し、品詞情報を付加する作業を形態素解析と呼ぶ。この形態素をタームと考えると、文はタームの繋がりであり、タームの組合せや出現数により解析を行うことができる。

ここで、 n 個のタームの繋がりを集計する方法を n -gram といい、2タームをバイグラム (bigram)、3タームをトライグラム (trigram) という。本研究では、3タームと2タームの共起を調べるため、以降はトライグラム、バイグラムと書く。トライグラムでは大まかに共起を調べ、バイグラムでその繋がりを調べることにより、統計処理を行う。

n -gram で見つけ出した共起がもつ意味や重要度は色々な指標が用いられるが、ここでは一般的な t 値と相互情報量 (Mutual Information) の MI 値を用いる。バイグラムで用いる相互情報量は、確率変数 X, Y の確率 $P(X), P(Y)$ を用いると次のように書ける。

$$MI = \log \frac{P(X, Y)}{P(X)P(Y)}$$

本研究では、これら t 値と MI 値を用いて共起の有意性を判断する。



図5 長期データのネットワーク図 (共起数 50 以上)

(1) 長期データの解析

長期データにおけるトライグラムの組み合わせは 129,870 であった。これにはハッシュタグ、絵文字などが多数含まれているが、これらを取り除いた時の組み合わせ数は共起数を 3 以上に限ると 3,320 になった。

ネットワーク図はタームをノード、共起の前後関係を方向、共起の多い少ないを線の太さで表示する。また、ネットワークはタームの隣接行列から作成するが、その隣接行列は $\{0,1\}$ により繋がりを示す。図 5 は共起数 50 以上の組合せのみ用いたグラフである。

図5から読み取れる情報は下関市、海響、角島大橋、調査捕鯨の4つである。これを考慮し、バイグラムで他の情報を確認する。

バイグラムでの探索では、ノードとなる形態素を定め、その前後5つの形態素を調べることにした。ただし、ノードとなる形態素は手作業で設定する。分析に用いたデータは、下関をキーワードで検索したこともあるが、図5では下関市が有意な情報であることがわかる。このため、初期ノードを下関としてバイグラムを調べた。このバブルチャートを図6に示す。なお、バブルの半径は共起数に対応する。また、これを表1に示す。

これらから、条約、大橋、店、シーモール、フグ、水族館、海峡、あるかぼーと、駅、独立、港、出港、徒歩、出身、中級の15タームのバイグラムを調べることにした。

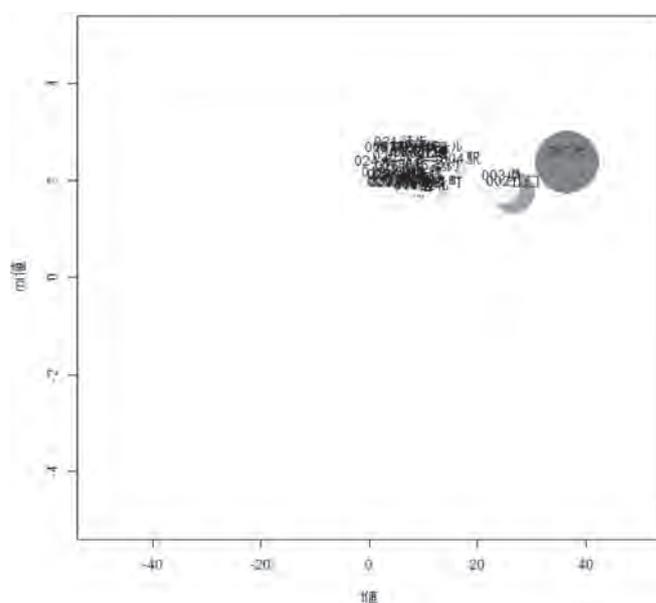


図6 下関をノードとするバブルチャート（共起数50以上）

表1 下関をノードとするバイグラムのt値とMI値（上位22件）

ラベル	t 値	mi 値	ラベル	t 値	mi 値
001. 市	36.612	2.361	012. 大橋	9.213	1.701
002. 山口	26.651	1.735	013. 市立	8.813	2.355
003. 県	24.450	1.855	014. 店	8.294	1.652
004. 駅	16.926	2.195	015. JR	7.876	1.801
005. 町	13.543	1.712	016. RedLine	6.811	2.428
006. 条約	11.704	2.061	017. 初	6.802	1.782
007. 港	11.275	2.316	018. IC	6.153	2.078
008. 豊北	9.751	1.672	019. 山陽本線	6.122	1.927
009. 神田	9.614	2.438	020. 水族館	6.025	1.739
010. シーモール	9.594	2.448	021. 徒歩	5.853	2.537
011. 出港	9.513	2.330	022. 出身	5.847	1.893

図7に条約をノードとしたバブルチャート、表2にt値とMI値を示す。これらから、このバイグラムは日清講和（下関）条約を締結したことを示していることがわかる。

同様に、大橋をノードとして、図8と表3を作成した。このバイグラムでは、角島大橋を「山口県下関市豊北町神田と同市豊北町角島間の海士ヶ瀬戸に架かる橋」と説明されている。

このようにバイグラムではt値およびMI値を用い、地域・観光情報で意味あるツイッターへの書き込みを調べた結果をまとめると角島大橋、日清講和条約とその歴史教育、調査捕鯨、海響館、フグ、地域情報（朝鮮初中等学校、シーモール専門店情報）、イベント情報（花火大会、あるかぼーとなど）の書き込みがあったことがわかった。

また、長期データの期間が7月から4月であることを考慮すると、暖かい季節には角島大橋、水族館（海響館）、花火大会、冬にはフグとそれに関わる事項（日清戦争、会場である春帆楼など）が話題になることは自然なことである。長期データでは、これらがツイッターで書き込まれていることが確認した。ただ、これ以上の地域・観光情報を見付け出すことはできなかった。それは、ツイッターの利用が20代や30代と行った年齢層であること、この年齢層のほとんどが都市部に出かけ、下関のような地域を観光することは少ないと考えるからである。地方都市を観光する年齢層がツイッター利用者よりも上の年齢層であることを考えると、本分析で有用な地域・観光情報が得られなかったことは不思議ではない。

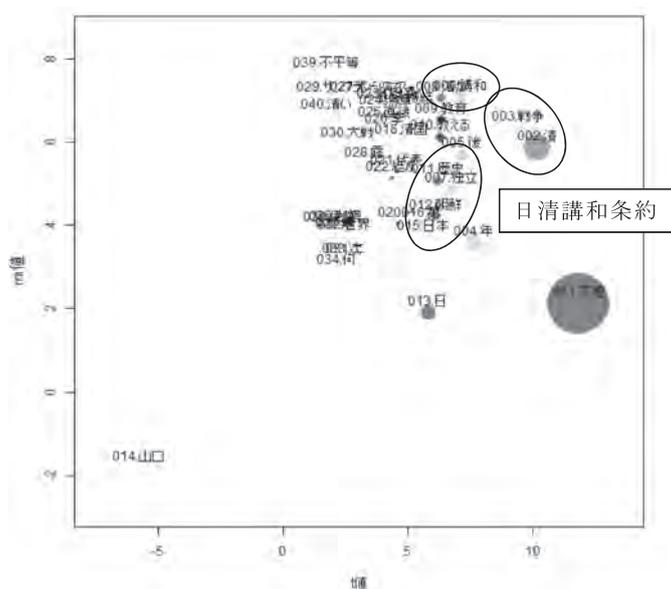


図7 条約をノードとするバブルチャート

表2 条約をノードとするバイグラムのt値とMI値（上位22件）

ラベル	t 値	mi 値	ラベル	t 値	mi 値
001. 下関	11.841	2.115	012. 朝鮮	6.132	4.214
002. 清	10.168	5.879	013. 日	5.821	1.907
003. 戦争	9.422	6.349	014. 山口	-5.785	-1.843
004. 年	7.617	3.591	015. 日本	5.620	3.717
005. 後	7.140	5.702	016. 第	5.553	4.025
006. 講和	7.089	7.079	017. 締結	4.955	6.788
007. 独立	6.759	4.858	018. 清国	4.722	6.020
008. 清か	6.355	7.051	019. 条	4.652	6.925
009. 教育	6.334	6.526	020. 一	4.600	4.028
010. 教える	6.311	6.123	021. 代表	4.571	5.301
011. 歴史	6.212	5.068	022. 結ぶ	4.346	5.144

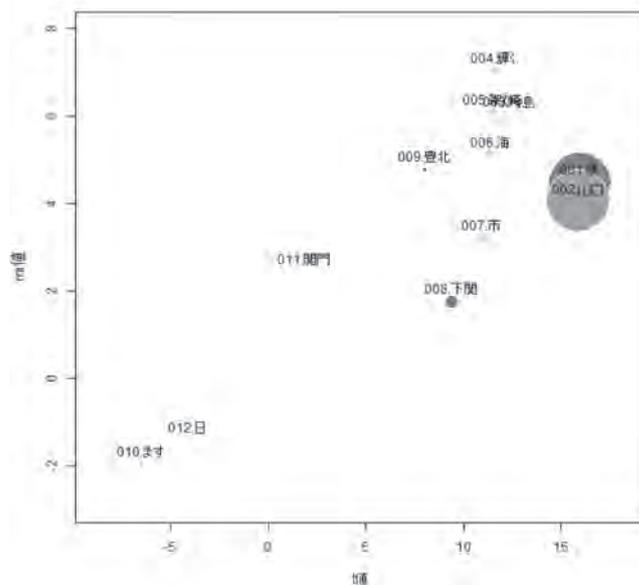


図8 大橋をノードとするバブルチャート

表3 大橋をノードとするバイグラムのt値とMI値

ラベル	t 値	mi 値	ラベル	t 値	mi 値
001. 県	15.958	4.486	007. 市	10.906	3.231
002. 山口	15.895	4.057	008. 下関	9.364	1.756
003. 角島	12.341	6.050	009. 豊北	8.006	4.786
004. 輝く	11.575	7.060	010. ます	-6.577	-1.979
005. 架かる	11.492	6.097	011. 関門	1.822	2.434
006. 海	11.410	5.123	012. 日	-4.121	-1.423

(2) 短期データの解析

短期データにおけるトライグラムの組み合わせは17,150であった。ハッシュタグ、絵文字などを取り除いた時の組合せ数は共起数3以上に限れば462である。

ここでは、主な結果のみを示す。共起数20以上のバブルチャートを図10に示す。トライグラムから見付かった情報はイベント、トラフグ、角島大橋である。長期データでも角島大橋はあったが、トラフグやイベントは見当たらなかった。これらを確認するため、バイグラムを用いる。長期データ同様、探索区間はノードの前後5形態素、初期ノードを下関とする。

表4に示すバイグラムのタームには駅、イベント、下関発のアイドルなど日常で使われるタームが目立つ。この短期データでは、表4にある開催、前売り、店、オープニング、とら、条約、アミバラ、直送、国産、魅力、シーモール、海鮮、希望の13タームについて調べる。ノードをとら、直送、国産とした場合、本場下関、国産、トラフグセット、直送というキーワードに繋がる。ここではとらをノードとした図11と表5を示す。

これ以外には日清講和条約、下関発アイドルPalmSugarが下関オープニングイベントに出演したという発言の多いユーザの書き込みが目立った。地域・観光情報に関しては、トラフグセット、日清講和条約以外は見当たらなかったが、下関発アイドルが注目されていることは新たな発見と言える。

これらをまとめると、短期データからはフグセット、下関駅オープニングイベント、日清講和条約、店情報などがツイッターで話題になっていたことがわかる。長期データにおける分析同様、ツイッター利用者の年齢層を考慮すると書き込まれている内容に芸能人、イベント、店舗情報が多いのは不思議ではない。これらツイッター利用者を地域や観光に呼び込むことを考えるのであれば、ここで示したキーワードは有効に機能する。ツイッターの書き込みをヒントにしたのではないが、淡路島ではロック、ヒップホップ、レゲエなどのアーティストを呼び、音楽フェスティバルを毎年開催し、イベント中は若者が集まっている。また、徳島市ではアニメ製作会社と連携したイベントを年数回開催している。

下関も特定の層をターゲットにした企画を立案する際、本研究で行った分析がヒントになるであろう。

表4 下関をノードとするバイグラムのt値とMI値

ラベル	t 値	mi 値	ラベル	t 値	mi 値
01. 市	14.280	1.935	12.PalmSugar	4.063	1.954
02. 駅	8.407	1.814	13. 事務所	3.895	2.954
03. 博多	7.705	1.879	14. 岬之	3.895	2.954
04. 上野	7.613	1.940	15. トラ	3.875	1.773
05. 開催	7.476	1.887	16. 条約	3.732	1.899
06. 神田	7.318	1.939	17. アミパラ	3.709	1.954
07. 前売り	7.038	1.954	18. J U N C H I	3.634	1.954
08. 希望	5.011	1.894	19. 波浪	3.558	1.954
09. オープニング	4.808	1.954	20. 詳細	3.557	1.664
10. 税理士	4.535	2.539	21. 直送	3.504	1.892
11. 店	4.370	1.618	22. ピン	3.480	1.954

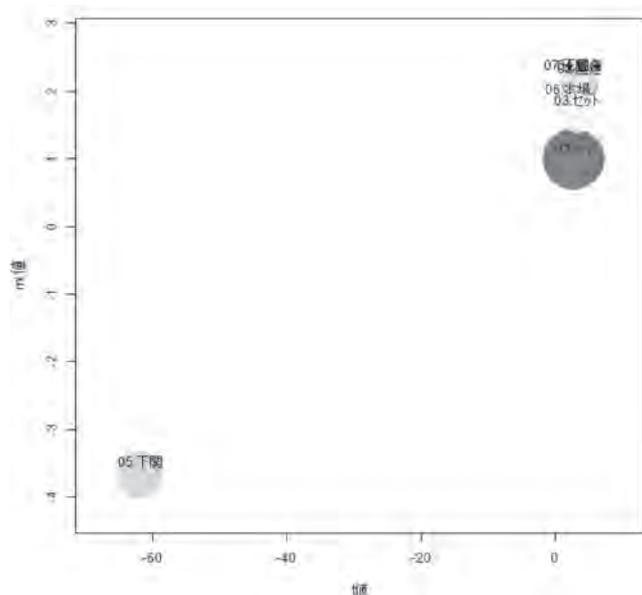


図11 とらえをノードとするバブルチャート

表5 とらえをノードとするバイグラムのt値とMI値

ラベル	t 値	mi 値	ラベル	t 値	mi 値
01. ふぐ	2.973	0.987	05. 下関	-61.925	-3.667
02. 直送	3.714	2.148	06. 本場	1.910	1.847
03. セット	3.308	1.689	07. 王様	1.753	2.209
04. 国産	3.676	2.209			

6. 結論

本研究では、インターネット上からの地域・観光情報収集し、収集したデータに対して分析を行い、その有用性を検討した。インターネット上に存在する膨大なデータの中から、地域や観光に関するデータを抽出し、観光推進・地域振興に役立つかどうかの検討を行った。分析対象としては、インターネット上における SNS である Twitter のツイートを洗濯した。Twitter は速報性・拡散性に秀でた SNS であり、様々な知見・情報を得ることが可能であると考えた。

分析は、長期データ（2013年7月～2014年7月）と短期データ（2014年3月5日）の2種類について行った。長期データについては、角島大橋・海響館・花火大会といった夏や暖かい季節の観光情報が多く含まれていた。また、フグ・日清講和条約などに関する発言も多く見られた。短期データについては、その時期のイベント情報や観光情報・フグに関する発言が多く見られた。どちらかといえば、長期データの方が地域・観光に関わるデータを多く取得することができた。

今回行った Web マイニングにおいては、一般的（市の観光ホームページ13）に掲載されていると思われる観光地（唐戸・カモンワフ・赤間神宮・巖流島・長府など）がほとんどキーワードとして検出されていない。これは、SNS を主に使用している 20代・30代においては、これらの観光地はあまり魅力的ではない可能性があると考えられる。逆に角島大橋・RedLine（ライブハウス）・PalmSugar（ご当地アイドル）などがキーワードとして多く検出された。これらから、今まで言われてきた、ふぐ・歴史だけではなく、別の視点による観光資源開発も必要になってくるのではないかと考えられる。SNS 等のインターネット上で話題になりやすい観光地は、若者向け・景観が良い（写真が投稿しやすい）場所が多いと考えられる。

今回の研究においては、「下関」を含むツイートを中心に分析を行ったが、「関門」「山口」「門司」などの関連する他のキーワードによる Web マイニングを行うことで別の知見が得られる可能性も考えられる。また、有名な観光地（北海道・沖縄・京都）などとの比較検討を行う必要性もあると思われる。

参考文献

- 1) 株式会社日本リサーチセンター、「メディアについての調査」、
<http://www.nrc.co.jp/report/131023.html>
- 2) 篠田 孝祐、榊 剛史、鳥海 不二夫、風間 一洋、栗原 聡、野田 五十樹、松尾 豊、「東日本大震災時における Twitter の活用状況とコミュニケーション構造の分析」、知能と情報、Vol.25, No.1, pp.598-608、2013 年
- 3) 内田 ゆず、荒木 健治、米山 淳、「ブログ記事からのオノマトペ用例文の自動抽出手法」、知能と情報、Vol. 24, No. 3 pp.811-820、2012 年
- 4) 株式会社 ICT 総研、「2013 年 SNS 利用動向に関する調査」、
<http://www.ictr.co.jp/report/20130530000039.html>(2014 年 6 月 1 日アクセス)
- 5) Twitter 社、「Twitter を利用している人」、
<https://business.twitter.com/ja/whos-twitter>(2014 年 6 月 1 日アクセス)
- 6) Twitter 社、「Interscope」、
<https://business.twitter.com/ja/success-stories/interscope>(2014 年 6 月 1 日アクセス)
- 7) Google 社、「Google Maps API」、<https://developers.google.com/maps/?hl=ja>(2014 年 6 月 1 日アクセス)
- 8) Amazon 社、「Amazon Product Advertising API」、
<https://affiliate.amazon.co.jp/gp/advertising/api/detail/main.html>(2014 年 6 月 1 日アクセス)
- 9) Twitter 社、「Changes coming in Version 1.1 of the Twitter API」、
<https://blog.twitter.com/2012/changes-coming-to-twitter-api>(2014 年 6 月 1 日アクセス)
- 10) Manabu Nii, Shota Miyake, Kazunobu Takahama, Atsuko Uchinuno, Reiko Sakashita, "Consideration about Utilizing Text Architecture for Making Feature Vectors in Classifying Nursing-Care Texts", IEEE International Conference on Systems, Man, and Cybernetics, pp.1817-1821, 2013
- 11) 中島 伸介、館村 純一、原 良憲、田中 克己、植村俊亮、「重要な blogger 発見を目的とした blog スレッド解析手法」、知能と情報、Vol.19, No.2, pp.156-166、2007 年
- 12) 村田 剛士、齋藤 皓太、「サイト・キーワードグラフを用いた Web ユーザの興味の抽出と視覚化」、知能と情報、Vol.18, No.5, pp.701-710、2006 年
- 13) 下関市役所、「しものせき観光ホームページ」、
<http://www.city.shimonoseki.yamaguchi.jp/kanko/model.html> (2014 年 6 月 1 日アクセス)