

位置情報付き Twitter データの 観光行動分析への有用性と限界

— 伊勢志摩地域における事例分析を通して —

桐 村 喬

〈要旨〉本稿では、代表的な位置情報付き SNS ログデータである Twitter データに注目し、どの程度の空間的スケールでの観光行動が行われているかを明らかにしながら、観光行動分析における Twitter データの有用性と限界について若干の考察を加える。分析に用いるのは、2012年4月から2015年3月までの3年度分の位置情報付きの Twitter データであり、ツイートに利用されたアプリの情報がモバイル機器向けのアプリであるものに絞り込んだものである。本研究における論点は主に2つである。

まず、Twitter データは近年様々な地理的分析において用いられているが、データのもつ地理的特性については十分に議論されていない。日本全国の Twitter データから最も投稿の多い地域を居住地として判定して、居住地別にユーザー数を集計すると、東京、大阪の二大都市圏への偏在が認められた。また、伊勢市において、観光客の実態調査に基づく居住地と比べると、Twitter ユーザーのほうが二大都市圏により偏在する傾向が認められた。観光行動の分析を行う場合は、居住地ごとに分析するか、あるいは居住地の偏りを補正したうえで分析する必要があると考えられた。

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

次に、Twitter データについては、2015年4月下旬以降、付与される位置情報が大きく変化し、従来学術研究に活用されてきたポイント単位のデータの大部分はチェックインサービスを通したものに限られるようになり、分析可能な空間単位は実質的にポイントから市区町村へと変化した。伊勢志摩地域を事例としてポイント単位のデータで観光行動分析を行った結果、伊勢市内の特定の地域間の行動と鳥羽市や志摩市との間の行動が確認された。伊勢志摩地域の場合、前者は2015年4月以降では詳細に分析することは難しいが、後者は今後も問題なく分析できるものと考えられる。位置情報の変更の結果、ポイント単位のデータは、全体を代表するものではなくなったため、今後の市内移動の分析においては、市区町村単位のデータも併用する必要があると判断された。

〈キーワード〉ビッグデータ、SNS、Twitter ユーザー居住地、三大都市圏、観光客流動

I はじめに

1. ビッグデータを用いた観光行動分析

従来、観光行動の実態に関する調査は、一定規模のサンプルへのアンケートによって行われてきた。例えば、国土交通省が行ってきた「全国旅行動態調査」は、公的機関が観光統計の作成のために行う代表的な調査である。また、当然のことながら、個別の研究者も観光客を対象に様々なアンケートを行ってきており、酒井ほか（1998）は、京都市を対象に約27,000件の調査票を配布する大規模な調査を実施している。アンケートを行うことで、実施機関や分析者が明らかにしようとする事柄を詳細に調査できる一方、通年の調査や継続的な調査は難しく、季節による観光行動の差や複数回の来訪者の行動など、アンケート調査のみでは把握しきれない事象も多い。

一方、2005年以降、国土交通省では、従来よりも包括的な観光統計の作成に着手し、「宿泊旅行統計調査」を2007年から実施するなど、観光統計の整備が進み、観光行動を分析するための調査資料の選択肢が広がっている。また、近年、情報通信技術の発達やGPS搭載の携帯電話の普及、ビッグデータの整

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

備・公開に伴い、様々な手段で観光客の行動を詳細に把握しようとする試みもみられる。

まず、矢部ほか（2010）は、GPS（Global Positioning System）を用いた人間や動物の行動調査に関する研究動向を整理したうえで、動物園来園者にGPSロガーを配布し、ミクロレベルでの観光行動調査におけるGPS利用とGIS（Geographic Information System）分析の有効性について検討している。GPSロガーの低価格化を受け、GPSを活用した観光行動分析は数多く行われるようになってきている（有馬，2010や杉本ほか，2013など）。アンケートと同様に、GPSによる調査は、研究者側が観光目的で来訪した対象者の行動を把握する手法である。これに対して、ソーシャルメディアなどを通して様々な人々が投稿・発信した情報である「自発的地理情報」（Volunteered Geographic Information）（Goodchild, 2007）から、観光に関するものを抽出し、人々の行動を把握しようとする試みもみられる。例えば倉田ほか（2015）は、写真共有サイトであるFlickrへの投稿データを利用して、観光地としての潜在的な可能性をもつ地域の抽出を試みている。また、鈴木（2016）は、Flickrに加えて、SNS（ソーシャルネットワーキングサービス）のTwitterに投稿された位置情報付きのログデータ（投稿データ）を利用して、観光地の格付けを行っている。また、同様に位置情報付きのTwitterのログデータを活用して、京都市（桐村，2013）や金沢市（渡辺，2016）などにおけるSNS利用者による観光行動パターンが分析されている。

SNS上に大量に蓄積された自発的地理情報は、いわゆるビッグデータとして観光行動の分析に活用されつつあるが、GPSによる調査と同様に利用可能性に関する議論も未だ並行して行われている（相，2014および中谷，2015）。また、TwitterやFlickrを運営するのは民間企業であり、倒産やサービスの停止・縮小、データの非公開化などが生じることもあり、研究者が永続的に利用できるとはいえない。例えばTwitterを運営する米国Twitter社は、2015年4月下旬に付与する位置情報を変更し、経緯度で示される位置情報よりも場所名による位置情報を優先して提供するようになった¹⁾。これによって、これまでのTwitterのログデータ（以下、Twitterデータと呼ぶ）の分析事例で行

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

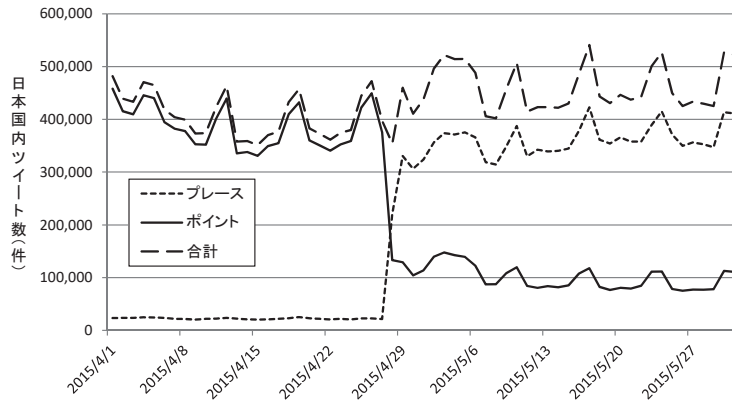
われてきたような観光地間の流動などの分析が可能なデータは大幅に減少しており、Twitter データが明らかにできる観光に関する事象は変化してきている。

2. Twitter データに付与される位置情報の変化とその影響

Streaming APIと呼ばれる Twitter 社が用意したツイート（投稿）データを取得するための仕組みを利用することで、Twitter のアカウントさえ所有していれば誰でも Twitter データを取得できる。筆者は、2012年2月から Streaming API を通して、日本全国を含む矩形の範囲内の位置情報が付与されているツイートを収集してきた。収集しているツイートの件数は、2017年2月までで約6億件である。2014年5月の日本国内のツイート数26億324万件²⁾ に対して、筆者が収集したのは1,110万件であることから、日本国内のツイートの0.4%を収集できていることがわかる。ツイート全体と比べて、収集できているツイートの割合は極めて少ないものの、Morstatter et al. (2013) によれば、Streaming APIを用いることで位置情報付きのツイートの大部分を収集できるとされており、Twitter 社による Streaming API の大幅な仕様変更がないと仮定すれば、筆者が収集しているデータについても、日本国内の位置情報を含むツイートの大部分が含まれていると考えられる。

Twitter 社による、ツイートに付与する位置情報の変更は、チェックインサービス³⁾ を提供する Foursquare Labs 社とのパートナーシップを契機として、2015年4月下旬ごろに行われた。Twitter 社は、経緯度の座標値で示された特定の位置（ポイント）ではなく、特定の場所（プレース）でのユーザーの体験に注目したいと考え、「プレースの情報が常にツイートに関連付けられるようになり、座標の情報は時々ツイートに関連付けられるように」なった⁴⁾。その結果、日本国内の位置情報を持つツイートのうちのポイントとプレースの比は、変更前におおむね15～19：1であったのに対し、変更後は0.2～0.3：1へと大きく変化した。日単位でのツイート数の推移をみると、4月27日・28日を境に、ポイント単位のツイート件数の激減と、プレース単位のツイート件数の急増を確認できる（第1図）。

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）



第1図 ポイント単位とブレース単位のツイート数の日単位推移 (2015年4月～5月)

資料：Twitter データ

ブレースとして付与されている位置情報の空間的なスケールは、日本の場合、主に市区町村である。2017年2月までのTwitter データに基づけば、ブレース単位の位置情報として付与されている市区町村には、2011年4月1日に愛知県西尾市に編入された一色町、吉良町、幡豆町が存在する一方、2010年4月1日に政令指定都市に移行した神奈川県相模原市に設置された緑区、中央区、南区が存在している。この間に市町村の合併や編入、区の設置はないことから、2010年4月2日～2011年3月31日の間のいずれかの時点の市区町村の情報がブレースとして利用されていると考えられる。

一方、付与される位置情報の変更によって、ポイント単位のTwitter データの性質はどのように変化したのだろうか。個々のツイートに含まれるツイートに利用されたアプリの情報からアプリ別の構成を集計すると、2015年3月の約1,276万件の日本国内でのポイント単位のツイートのうち57.8%が iPhone 用の Twitter アプリから投稿され、次いで Android 用の Twitter アプリから 26.1% が投稿されていた (第1表)。Foursquare を通したツイートは全体の 6.2% であり、3 番目に多かった。位置情報の変更後の2015年5月では、約308万件の日本国内でのポイント単位のツイートのうち、最多は27.8%の Foursquare を通したものであった。iPhone 用、Android 用の Twitter アプリ

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

第 1 表 2015 年 3 月・5 月の主要アプリ別ツイート数
（日本国内および伊勢志摩地域）

		日本国内		うち伊勢志摩地域	
		2015年3月	2015年5月	2015年3月	2015年5月
iPhone用Twitterアプリ	件数	7,373,807	634,315	15,440	1,291
	割合	57.8%	20.5%	61.2%	25.9%
Android用Twitterアプリ	件数	3,329,294	628,986	7,457	1,622
	割合	26.1%	20.4%	29.6%	32.6%
Foursquare	件数	792,448	858,386	1,025	888
	割合	6.2%	27.8%	4.1%	17.8%
その他のアプリ ・サービス経由	件数	1,264,855	965,884	1,312	1,174
	割合	9.9%	31.3%	5.2%	23.6%
全体	件数	12,760,404	3,087,571	25,234	4,975
	割合	100.0%	100.0%	100.0%	100.0%

資料：Twitter データ

からのツイートは合わせて40.9%に過ぎず、割合は半減している。Foursquare を通したツイートの割合は大きく上昇しているが、2015年3月は約79万件、2015年5月は約86万件であり、大幅な増加は認められない。iPhone および Android というスマートフォン上の Twitter アプリからのツイートの減少により、Foursquare の割合が高まっていることになる。伊勢志摩地域⁵⁾ に限定すれば、2015年3月は類似した傾向を示すものの、2015年5月の Foursquare が占める割合は日本国内全体のものより低い傾向にあった。Foursquare 上で示すことのできるポイント単位の位置情報の網羅性や Foursquare の利用率が地域によって異なると考えられ、ポイントからプレスへの付与される位置情報の変更の影響は、地域によって異なる可能性がある。

従来、ツイートに含まれる経緯度単位の位置情報と、Twitter ユーザー（以下、単にユーザーと呼ぶ）の ID に基づいて、ユーザーの行動が分析されてきた。付与される位置情報が市区町村単位中心に変更されることで、ユーザーの行動を分析できる空間的スケールは大きくなってしまい、2015年4月下旬までのデータのように市区町村よりも小さい空間的スケールで行動を詳細に分析することは難しくなっている。また、伊勢志摩地域の例のように、Foursquare などのチェックインサービスの位置情報が付与されたツイートが少ない地域の場合、把握できるユーザーの行動はより限定されたものになることが予想される。

2. 研究目的・分析データ

本稿では、代表的な位置情報付き SNS ログデータである Twitter データに注目し、ポイント単位の位置情報が付与されたツイートが多数であった2015年4月以前のデータを用いて、どの程度の空間的スケールでの観光行動が行われているかを明らかにしながら、観光行動分析における Twitter データの有用性と限界について若干の考察を加える。観光行動に関する事例分析の対象地域は伊勢志摩地域である。

分析に用いるのは、2012年4月から2015年3月までの3年度分の位置情報付きの Twitter データであり、このうち、ツイートに利用されたアプリの情報がモバイル機器向けのアプリ⁶⁾であるものを分析対象とする。このように絞り込むのは、ボットと呼ばれる機械的な投稿を行うアプリや特定のアプリによって付与された偽の位置情報を含むツイートを除外するためであり、これによって、ツイートから得られる位置情報の精度は一定程度向上する。このように絞り込んだデータのうち、II章では、日本国内の位置情報を含むものすべてを利用し、III章では伊勢志摩地域の位置情報を含むもののみを利用する。

II Twitter ユーザーの居住地判定

1. 居住地の判定方法

観光地における来訪者の行動は、来訪者自身の居住地によって異なる。一般的には、居住地と観光地の距離によって滞在期間が変化し、それによって観光行動も異なってくる。そこで、ユーザーの行動を分析する前に、ユーザーの居住地を判定する。

ユーザーの居住地を判定する最も簡便な方法は、最も多く投稿した地域を居住地とするものである。ただし、その地域をどのような空間単位で捉えるかによって、居住地は変動する。例えば、ツイートは、自宅やその近接地域だけでなく、従業地への通勤途上や従業地、帰路の乗換駅など、日常生活における様々な地点で行われる。そのため、市区町村のような単位ではなく、生活圏を含めた都市圏単位で居住地の判定を行う必要がある。代表的な都市圏の定義として、総務省による大都市圏・都市圏⁷⁾と、金本らによる大都市雇用圏

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

(MEA)⁸⁾とがあり、前者はより広域的であり、圏域の数も少ないことから、ここでは2010年の国勢調査結果に基づく後者を用いる。MEAが設定されていない地域については、2010年の国勢調査時点（10月1日）の市町村を用いる。

分析対象のユーザー数は約303万であり、彼らのツイート総数は約2億6,585万件である。居住地と判定されたMEA別にユーザー数を集計すると、最も多い東京MEAが924,236ユーザーであり、大阪MEAの284,294、名古屋市・小牧市MEAの106,812と続いている（第2表）。ユーザー数の多い上位10のMEAのうち、国勢調査による2010年の常住人口1,000人当たりのユーザー数が日本全国の23.7を上回るのは、東京MEAと京都市MEAであり、名古屋市・小牧市MEA、神戸市MEAなどは全国の値を下回っている。大阪MEAは日本全国と同等であるものの、ユーザーの居住地は全体として二大都市圏に偏った分布を示している。

第2表 居住 Twitter ユーザー数上位10の大都市雇用圏（MEA）

大都市雇用圏 (MEA)	居住Twitter ユーザー数	常住人口1,000人当たり 居住Twitterユーザー数
東京	924,236	26.5
大阪	284,294	23.2
名古屋市・小牧市	106,812	19.5
京都市	72,636	27.1
福岡市	53,890	21.6
札幌市・小樽市	50,784	21.7
神戸市	43,853	18.0
仙台市	37,146	23.6
前橋市・高崎市・伊勢崎市	25,769	17.7
広島市	25,285	17.9
日本全国	3,037,312	23.7

資料：Twitter データ、2010年大都市雇用圏統計データ

2. 伊勢市における観光客の居住地別構成と Twitter ユーザーの居住地別構成

伊勢志摩地域のうちで、観光客実態調査が行われている伊勢市に注目して、実際の観光客の居住地別構成とユーザーの居住地別構成を比較し、居住地分布の偏りについて検討する。

伊勢市のウェブサイトの詳細な資料が公開されている2015年の観光客実態調

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

査の結果⁹⁾ から、都道府県別に集計された伊勢市以外に居住する観光客の居住地別構成をみると、愛知県（19.8%）、三重県（14.7%）、大阪府（12.4%）、東京都（7.1%）の順に続いている。一方、2012～2014年度中に伊勢市で投稿したユーザーの居住地別構成をみると、伊勢市MEAに居住するユーザー（3,515）を除いた21,379ユーザーのうち、東京MEAが最も多い5,348（25.0%）、大阪MEAが次いで4,095（19.2%）、名古屋市・小牧市MEAが3,213（15.0%）となっており、集計空間単位の違いを考慮しても、東京MEAや大阪MEAへの偏りが大きいものと考えられる（第3表）。

第3表 伊勢市に訪問経験のある Twitter ユーザーの
主要居住大都市雇用圏（MEA）

大都市雇用圏(MEA)	居住Twitter ユーザー数	割合
東京	5,348	25.0%
大阪	4,095	19.2%
名古屋市・小牧市	3,213	15.0%
津市	1,624	7.6%
京都市	1,141	5.3%
その他	5,958	27.9%
伊勢市以外の合計	21,379	100.0%
伊勢市	3,515	-

資料：Twitter データ

このように、ユーザーの居住地は、常住人口に対する比率、実際の観光客の居住地別構成のそれぞれと比べても、二大都市圏への一定の偏りが存在していることは明らかであり、その点を考慮しながら分析を進める必要がある。なお、本来であれば、このような居住地の偏りを補正する方法を開発する必要があるが、これについては別稿に譲り、MEA別の分析を行うことで居住地の偏りの影響を避けることにする。

Ⅲ 伊勢志摩地域における Twitter ユーザーの観光行動

1. Twitter ユーザーの基本的動向

伊勢志摩地域における分析対象のツイート総数は423,475件であり、ユーザーの総数は24,894である。日本国内が居住地と判定されたのは、このうちの

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

24,808 ユーザーであり¹⁰⁾、MEA 別にみれば東京MEAが最も多い5,348 (21.6%)、次いで大阪MEAが4,095 (16.5%)、伊勢市MEAが3,515 (14.2%)と続いている(第4表)。伊勢志摩地域のうち明和町と大紀町を含む津市MEAは1,624 (6.5%)、非MEAである志摩市は736 (3.0%)となっており、伊勢志摩地域およびその周辺地域を合わせても、地域外からのユーザーのほうが多い状況にある。ただし、ツイート数自体は、伊勢市MEAが過半数を占めており(53.7%)、3分の2程度が地域内に居住するユーザーによる投稿と考えられる。

第4表 伊勢志摩地域に訪問経験のある Twitter ユーザーによる
主要居住大都市雇用圏(MEA)別ユーザー数・ツイート数

大都市雇用圏(MEA)	Twitterユーザー数		ツイート数	
	件数	割合	件数	割合
東京	5,348	21.6%	29,971	7.1%
大阪	4,095	16.5%	23,340	5.5%
伊勢市	3,515	14.2%	227,534	53.7%
名古屋市・小牧市	3,213	13.0%	20,261	4.8%
津市	1,624	6.5%	39,800	9.4%
京都市	1,141	4.6%	5,594	1.3%
志摩市*	736	3.0%	49,692	11.7%
四日市市	697	2.8%	4,573	1.1%
神戸市	388	1.6%	2,378	0.6%
岐阜市	233	0.9%	1,180	0.3%
浜松市	193	0.8%	698	0.2%
非MEAを含む全体	24,808	100.0%	423,475	100.0%

*志摩市は大都市雇用圏単位ではなく市単位である。

資料：Twitter データ

本章では、三大都市圏(東京、名古屋、京阪神)の中心となる5つのMEAに居住するユーザーに注目して、伊勢志摩地域における彼らの行動を観光行動と見なして分析する。対象とする5つのMEAは東京MEA、名古屋市・小牧市MEA、京都市MEA、大阪MEA、神戸市MEAであり、これ以降ではそれぞれ東京、名古屋、京都、大阪、神戸と省略して表記し、これらをまとめて三大都市圏と呼ぶ。これらの地域からの来訪者の多くは通勤・通学者ではなく、ビジネス需要もあるとは考えられるものの、広い意味では観光客と考えられる。三大都市圏に居住するユーザーは合計で14,185であり、伊勢志摩地域を訪問したユーザー全体の57.2%を占める。

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

2. Twitter ユーザーの観光行動の把握手順

三大都市圏からのユーザーの観光行動を把握するために、ツイートの位置情報から主要な訪問地を特定したうえで、ユーザーの伊勢志摩地域外からの流入と伊勢志摩地域内での観光行動を把握する。手順は以下のとおりである。

まず、三大都市圏に居住するユーザーのツイートについて、ArcGIS のカーネル密度推定ツールによって、バンド幅 500m 以内のツイート密度を 100m 四方のメッシュ単位で求め、1 平方キロメートルあたり 1,000 件以上のメッシュを抽出した。メッシュが連担する場合は 1 つの地域としてまとめ、伊勢志摩地域内でツイートの多い 25 の地域を抽出した。さらに、この 25 地域の外縁から半径 100m のバッファを生成し、各地域内への 3 年間の訪問ユーザー数をユーザー ID 単位で集計した。25 地域のうち、訪問ユーザー数が 200 以上である 11 地域を主要訪問地として分析対象とする（第 5 表）。

第 5 表 分析対象とする 11 の主要訪問地

訪問地名称	訪問ユーザー数	所在地
内宮周辺	7,121	伊勢市
外宮周辺	4,664	伊勢市
鳥羽	2,263	鳥羽市
二見浦	1,198	伊勢市
五十鈴川	880	伊勢市
賢島	478	志摩市
志摩スペイン村	402	志摩市
サンアリーナ	376	伊勢市
鵜方	300	志摩市
相差	290	鳥羽市
鳥羽展望台	235	鳥羽市

資料：Twitter データ

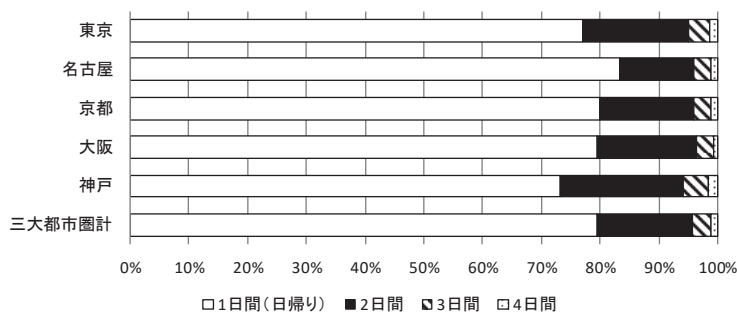
次に、ユーザーの伊勢志摩地域における観光行動については、滞在期間中に毎日投稿が行われるわけではなく、また、同一ユーザーによる複数回の訪問もみられることから、旅行の単位を、ツイートの間隔に基づいて設定する必要がある。ここでは、週末を中心とする日数の旅行を想定し、最初のツイートから次のツイートまで、その日を含んで 4 日以内の間隔である限りは、1 つの旅行と判断する。三大都市圏に居住するユーザーによる延べ 18,422 件の旅行について、最初のツイートの日を含めた滞在日数別にユーザー数を集計すると、1 日

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

間（日帰り）が大部分を占めている。以降の分析では、延べ旅行総数の大部分を占める1日間から、1%以上を占めている4日間までの計18,119件の旅行を分析対象とする。

3. Twitter ユーザーの居住地別にみた滞在日数と訪問回数

分析対象のユーザーによる旅行を、居住地および滞在日数別にみれば、いずれの地域でも日帰りの割合が最も高く、少なくとも70%以上を占めている。日帰りの割合が最も高いのは名古屋の83.3%であり、三大都市圏のうちでは東京と神戸で日帰りの割合がやや低い傾向にある（第2図）。日帰りの割合が相対的に低い地域では、2日間以上の滞在日数の旅行が多くなっている。三大都市圏のうち、名古屋は伊勢志摩地域に最も近く、大阪と京都はそれに次いで近い。神戸は、阪神なんば線の開通により伊勢志摩地域へのアクセスが改善されたとはいえ、大阪よりもさらに遠方にあり、相対的に長期の滞在が多くなっているものと考えられる。

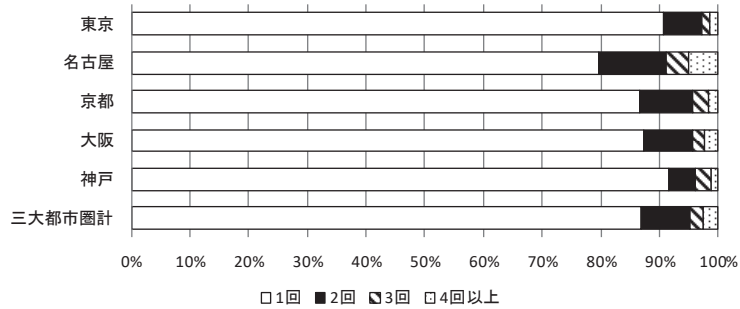


第2図 伊勢志摩地域における Twitter ユーザーの居住地別滞在日数

資料：Twitter データ

ユーザーごとの訪問回数については、1回限りのユーザーが大部分を占めているものの、地理的、時間的に近接する地域ほど複数回の訪問の割合が高い傾向を示している（第3図）。すなわち、滞在日数が短い地域ほど、訪問回数が多いことになる。

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）



第3図 伊勢志摩地域におけるTwitterユーザーの居住地別訪問回数

資料：Twitter データ

4. Twitter ユーザーの居住地別にみた観光行動

まず、ユーザーの居住地別に、主要訪問地別の延べ訪問回数を求め、居住地別の延べ旅行総数に対する割合を算出し、三大都市圏全体での割合に対する居住地ごとの特化係数を求めた（第6表）。東京からは、外宮周辺、内宮周辺、二見浦、鳥羽、五十鈴川への訪問が多い一方、大阪からは志摩スペイン村、賢島、鵜方、相差への訪問が多いことが確認された。京都は鳥羽展望台、相差、二見浦への訪問が多く、神戸は大阪からの訪問が多い訪問地と京都からの訪問が多い訪問地を合わせたようなパターンである。一方、名古屋からはサンアリーナへの訪問が多いのみであり、他の主要訪問地への訪問は低調である。東

第6表 伊勢志摩地域におけるTwitterユーザーの訪問地の特徴

	居住地別の延べ旅行数に対する割合						特化係数*				
	東京	名古屋	京都	大阪	神戸	三大都市圏全体	東京	名古屋	京都	大阪	神戸
内宮周辺	55.9%	36.7%	38.8%	39.7%	45.3%	44.6%	125.4	82.2	87.1	89.1	101.6
外宮周辺	56.7%	30.1%	25.4%	27.7%	35.3%	38.4%	147.6	78.4	66.2	72.1	91.9
鳥羽	17.7%	10.5%	14.7%	15.1%	18.2%	14.8%	119.5	70.9	99.1	101.8	123.0
二見浦	8.7%	4.9%	9.3%	6.5%	10.7%	7.2%	121.2	69.0	130.5	90.9	149.1
五十鈴川	7.5%	6.3%	5.1%	6.6%	6.9%	6.7%	111.8	93.0	76.2	98.3	102.3
賢島	2.8%	2.1%	2.2%	4.5%	3.6%	3.1%	90.5	69.0	72.2	146.6	116.1
志摩スペイン村	1.2%	2.1%	2.8%	4.8%	3.8%	2.7%	45.8	80.1	104.3	179.8	141.7
サンアリーナ	2.9%	3.5%	2.5%	1.7%	2.2%	2.7%	109.1	131.7	94.0	62.3	83.7
鵜方	1.8%	2.5%	1.3%	3.5%	2.9%	2.4%	72.9	102.9	52.4	141.4	118.0
相差	1.9%	1.3%	2.6%	2.4%	2.7%	2.0%	97.2	65.2	134.2	123.9	135.7
鳥羽展望台	1.2%	1.4%	2.7%	1.6%	2.0%	1.5%	79.4	90.9	180.8	109.1	133.5
延べ旅行数	6,390	4,920	1,403	5,259	450	18,422					

*特化係数は、居住地別割合を三大都市圏全体の割合で除して100を乗じた値。100以上であれば、全体の分布よりも偏っていることになる。

資料：Twitter データ

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

京からは伊勢市および鳥羽市への訪問傾向が顕著であり、京阪神からは鳥羽市および志摩市への訪問が中心であることがわかる。

次に、主要訪問地間あるいは居住地と訪問地間のユーザーの流動を、旅行単位で検討する。主要訪問地のうちで最初に投稿する訪問地、すなわち入口となる訪問地は、いずれの居住地からも内宮周辺と外宮周辺が大部分を占め、鳥羽がそれらに続いている（第7表）。ただし、東京と神戸を除けば、主要訪問地

第7表 伊勢志摩地域における Twitter ユーザーの最初の訪問地

		居住地				
		東京	名古屋	京都	大阪	神戸
最初の訪問地	内宮周辺	28.2%	26.6%	25.7%	25.9%	29.5%
	外宮周辺	37.1%	19.6%	19.2%	19.4%	22.4%
	鳥羽	8.8%	6.9%	9.7%	8.9%	9.2%
	二見浦	3.2%	2.6%	5.4%	3.2%	7.6%
	五十鈴川	1.6%	2.8%	1.7%	2.3%	1.8%
	賢島	1.3%	1.2%	1.3%	2.2%	2.1%
	志摩スペイン村	0.5%	1.5%	1.9%	2.8%	2.5%
	サンアリーナ	1.2%	2.1%	1.3%	0.8%	1.4%
	鵜方	0.8%	0.8%	0.7%	1.6%	0.7%
	相差	0.7%	0.9%	0.9%	1.3%	1.8%
	鳥羽展望台	0.5%	0.8%	1.4%	1.1%	1.1%
	その他の地域	15.9%	34.2%	30.8%	30.5%	19.9%
延べ旅行数		6,199	4,753	1,378	5,164	437

資料：Twitter データ

以外の伊勢志摩地域内でのみ投稿した旅行が3割以上存在している点に注意が必要である。鉄道による訪問が中心と考えられる東京の場合、近鉄のターミナル駅が含まれる外宮周辺を最初の到着地とする割合が、他の居住地と比べて高い傾向にあり、利用交通手段の特徴が表れている。一方、最後に投稿する出口となる訪問地については、いずれの居住地でも内宮周辺が最多であり、外宮周辺、鳥羽と続いている（第8表）。

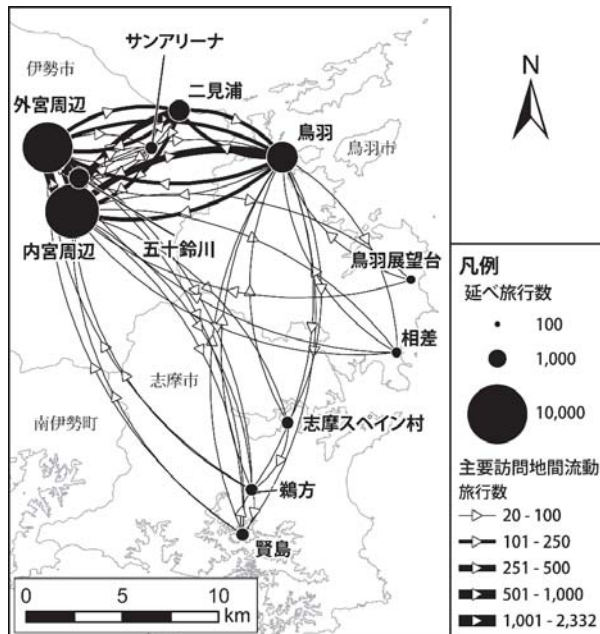
最後に、主要訪問地間の観光行動は第4図のようになっている。すべての行動パターンについて検討することは困難であるため、いずれかの方向で延べ100件以上の旅行が確認できる主要訪問地間について、方向別に延べ旅行数をまとめた（第9表）。居住地ごとの差はそれほど顕著ではなかったため、表には示していない。外宮周辺から内宮周辺への流動がその逆方向の2.7倍あり、

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

第 8 表 伊勢志摩地域における Twitter ユーザーの最後の訪問地

		居住地				
		東京	名古屋	京都	大阪	神戸
最後の訪問地	内宮周辺	38.7%	29.5%	29.8%	29.8%	35.9%
	外宮周辺	23.1%	15.4%	12.0%	13.9%	16.2%
	鳥羽	10.7%	7.5%	10.0%	9.5%	11.2%
	二見浦	3.4%	2.9%	6.2%	3.5%	5.7%
	五十鈴川	2.4%	3.0%	3.2%	2.7%	2.5%
	賢島	1.5%	1.3%	1.5%	2.3%	1.8%
	志摩スペイン村	0.8%	1.5%	2.2%	2.7%	1.8%
	サンアリーナ	1.3%	2.2%	1.0%	0.8%	0.9%
	鵜方	0.8%	0.9%	0.9%	1.9%	1.4%
	相差	0.8%	0.8%	1.1%	1.4%	1.6%
	鳥羽展望台	0.5%	0.9%	1.5%	1.0%	0.9%
	その他の地域	15.9%	34.2%	30.8%	30.5%	19.9%
延べ旅行数		6,199	4,753	1,378	5,164	437

資料：Twitter データ



第 4 図 伊勢志摩地域における Twitter ユーザーの主要訪問地別延べ旅行数と流動

資料：Twitter データ

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

外宮参拝後ないしは鉄道で伊勢市入りして内宮周辺に移動する流れが多数を占めていることがわかる。また、内宮周辺からは、鳥羽、二見浦、五十鈴川への流動も大きく、それらの逆方向の流動を上回っており、内宮周辺の訪問のうちに、周辺の観光地を訪問する傾向にあると考えられ、外宮周辺および内宮周辺が観光行動の軸になっている。

第 9 表 延べ旅行数 100 件以上の主要訪問地間の方向別旅行数

主要訪問地	方向と旅行数	主要訪問地
内宮周辺	→ 860 → ← 2,332 ←	外宮周辺
内宮周辺	→ 373 → ← 250 ←	鳥羽
外宮周辺	→ 225 → ← 194 ←	鳥羽
内宮周辺	→ 271 → ← 188 ←	二見浦
外宮周辺	→ 179 → ← 179 ←	二見浦
鳥羽	→ 101 → ← 133 ←	二見浦
内宮周辺	→ 347 → ← 217 ←	五十鈴川
外宮周辺	→ 179 → ← 171 ←	五十鈴川

資料：Twitter データ

IV 観光行動分析における Twitter データの有用性と限界

1. 伊勢志摩地域における Twitter ユーザー行動の特徴

Twitter データの分析から、伊勢志摩地域におけるユーザーの観光行動の特徴として、東京・大阪に居住するユーザーが多いこと、伊勢志摩地域からの地理的・時間的距離が近いほど滞在日数が短くなり、訪問回数が増加すること、東京からは内宮・外宮を中心とする伊勢市と鳥羽市への訪問が多い一方、京阪神からは鳥羽市と志摩市への訪問が多いこと、外宮周辺から内宮周辺、内宮周辺から各観光地への移動という基本的行動パターンが形成されていることが確認された。居住地によるユーザーの行動パターンの差異は、居住地から伊勢志摩地域までの地理的・時間的距離の差によって説明されるというよりは、むしろ、ユーザーによる訪問先の選好の差異によって説明できよう。三大都市圏の

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

うちで最も遠い東京のユーザーは、外宮周辺と内宮周辺に主な関心を抱いて伊勢志摩地域を訪問すると考えられ、比較的近接し、複数回の訪問が可能である名古屋や京阪神からは、内宮・外宮に限らない様々な伊勢志摩地域内の観光地を訪問する傾向にあるといえる。志摩市については、京阪神、特に大阪からの訪問が多い傾向にある一方、より近い名古屋からの訪問はそれほど多くはなかった。例えばナガシマリゾートのように、名古屋の周辺に存在し、伊勢志摩地域よりも近い場所にある、志摩市と類似した観光資源を抱える地域との競合関係によって、名古屋からの訪問が少ない可能性が考えられる。

Twitter データの分析からは、三大都市圏のユーザーにおける伊勢志摩地域での観光行動の基本的パターンを明らかにできた。Twitter データにはユーザーの属性情報が存在しないが、この分析結果を活用して、多くのユーザーが訪問する外宮周辺と内宮周辺を中心としてアンケート調査を詳細に実施すれば、より多くの有用な情報を効率的に集めることができるだろう。また、ここでは検討しなかったが、Twitter データからは、複数回訪問したユーザーのそれぞれの訪問時の訪問場所も把握できるので、いわゆるリピーターによる行動の特徴や変化を明らかにすることもできる。どちらか一方のみに頼るのではなく、Twitter データの分析とアンケートによる分析を並行して実施することで、より詳細に観光行動を解明することができると考えられる。

2. Twitter ユーザーの地理的偏在

伊勢市における事例分析から、ユーザーの居住地は常住人口および観光客の実態のいずれにおいても、二大都市圏への一定の偏りが存在することが確認された。例えばユーザーの居住地別に観光行動の分析を行う場合は、このような偏りによる分析結果への影響はそれほど大きくはない。III章3節・4節での分析においても、居住地ごとに独立して割合や特化係数などを求めることで、ユーザーの居住地の偏りの影響を受けないように留意した。一方で、訪問地ごとに、どの地域に居住するユーザーが多いかという視点から居住地別の構成比を求めるような分析は、この偏りの影響を受けやすい。居住地ごとの総数ではなく、全体の総数に対する居住地別の割合や内訳について言及する場合は、観

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

光客の実態調査の結果などに基づいて、補正を行う必要がある。補正する基準となるデータがない場合は、常住人口や、ユーザーの居住地分布とより相関関係の強い別の指標を用いて補正する必要がある。

3. 分析可能な空間的スケールの変化の影響

今回の分析では、ポイント単位の位置情報が付与された Twitter データを利用したものの、2015年4月以降のデータの大部分はプレース単位、主に市区町村単位のデータである。伊勢志摩地域であれば、外宮周辺、内宮周辺、五十鈴川、二見浦など、伊勢市内の主要訪問地間における観光行動の分析は、市区町村単位の位置情報しか付与されていないデータでは分析できないことになる。また、伊勢志摩地域の場合、Foursquare を通して投稿されるデータは全国平均よりも少なく、ポイント単位のデータの減少の影響は大きい。加えて、いわゆる「平成の大合併」により、非大都市圏を中心に市町村合併が進んでおり、そのような地域では、Twitter データでの分析可能な空間的スケールは大きくなってしまっている。

一方、市区町村単位のデータでも、東京からの伊勢市・鳥羽市志向と、京阪神からの鳥羽市・志摩市志向のような居住地による訪問先の差異や、伊勢市と鳥羽市間の流動など、伊勢志摩地域でも確認された市区町村間の観光行動のパターンは把握できる。また、政令指定都市や東京23区であれば、大都市でもある程度の市内流動の分析が可能であるが、松山市や鹿児島市のような人口50万人程度の都市内部の流動は把握できない。加えて、本稿で用いたような居住地の判定方法であれば、市区町村単位のデータでも影響はほとんどなく、市区町村単位の Twitter データは、広域的な観光行動の分析には現在も適していると考えられる。

以上のことから、2015年4月下旬以降について、一定程度の広域を対象とし、市区町村単位で観光行動分析を進める場合は市区町村単位の Twitter データを利用すればよいと考えられる。しかし、ポイント単位の Twitter データを活用して市区町村内の観光行動についても検討する場合は、市区町村単位の Twitter データで基本的な動向を確認したうえで、ポイン

位置情報付き Twitter データの観光行動分析への有用性と限界（桐村）

ト単位での分析を進めるほうがよいだろう。2015年4月下旬以降のポイント単位の Twitter データは、どちらかといえば Foursquare などのチェックインサービスの利用者によるツイートが主であり、ポイント単位の Twitter データのみを利用してユーザーの行動を代表させることは難しいためである。また、伊勢志摩地域のように、Foursquare を通したツイートが少ない地域では、より限られたユーザーの行動しか把握できないことになり、チェックインサービスの利用状況の地域差についても考慮する必要がある。

V おわりに

本稿では、代表的な位置情報付き SNS ログデータである Twitter データに注目し、伊勢志摩地域を事例とした基礎的な事例分析を通じて、観光行動分析における Twitter データの有用性と限界の一端を示した。伊勢志摩地域における 2012年4月から2015年3月までの Twitter データの分析の結果、訪問ユーザーの二大都市圏への居住地の偏りや、居住地による伊勢志摩地域内の訪問先、行動パターンの差異が明らかになった。本稿における観光行動分析に対する Twitter データの有用性と限界に関する論点と問題の改善策は次の2点である。第一に、ユーザーの居住地の偏りについては、実際に二大都市圏に一定の偏りが確認されたことから、全体をそのまま母数とはせず居住地ごとに分析するか、居住地の偏りを人口などによって補正したうえで分析する必要があることが示された。第二に、付与される位置情報の変更によって、ポイント単位のデータで分析できるのは、主にチェックインサービスを利用する一部のユーザーの行動に限られるようになったため、これまでのようにポイント単位のデータのみで観光行動を分析するのではなく、市区町村単位のデータを併用して全体の観光行動の動向を把握しつつ、ポイント単位での分析を行う必要性が示された。伊勢志摩地域のように Foursquare を通したツイートが少ない地域については、より限られたユーザーによるツイートに基づく分析になってしまうため、十分に留意する必要がある。

Twitter データは、従来の研究でも示されてきたように、観光行動分析にとって有用かつ重要な意味を持つデータである。しかし、ユーザーの居住地の

位置情報付き Twitter データの観光行動分析への有用性と限界 (桐村)

偏りや、付与される位置情報の変更によって、用途や分析手法に一定の制限が生じてしまっている。紙幅の都合もあり、本稿は基礎的な分析かつ浅い議論となってしまったが、居住地の偏りの補正方法の開発や付与される位置情報の変化とその影響についての詳細な分析が今後必要となろう。

謝辞

本研究は JSPS 科研費 15H03344 の助成を受けたものである。

文献

- 相 尚寿 (2014). 観光研究への位置情報ビッグデータ展開の可能性. 観光科学研究, 7, 11-19.
- 有馬貴之 (2010). 動物園来園者の空間利用とその特性－上野動物園と多摩動物公園の比較－. 地理学評論, 83(4), 353-374.
- 桐村 喬 (2013). 位置情報付きツイッター投稿データにみるユーザー行動の基本的特徴－観光行動分析への利用可能性－. 地理情報システム学会講演論文集, 22, CD-ROM.
- 倉田陽平・相 尚寿・真田 風 (2015). 写真共有サイト投稿データを利用した新たな観光マップの構築. 観光科学研究, 8, 151-154.
- 酒井 弘・東 徹・西井和夫・中村嘉次 (1998). 京都市における観光交通行動の実態調査手法および分析. 土木計画学研究・講演集, 21(2), 591-594.
- 杉本興運・岡野祐弥・菊地俊夫 (2013). レンタサイクル利用による観光回遊行動の実態－長野県安曇野市におけるGPS・GIS支援による調査とデータ解析－. 観光研究, 24(2), 15-27.
- 鈴木英之 (2016). 複数の位置情報付きソーシャルメディア等を用いた観光地の格付け. 地理情報システム学会講演論文集, 25, CD-ROM.
- 中谷友樹 (2015). 外国人旅行者の行動空間に関する地理的可視化－京都市を対象としたTwitterおよびGPS調査資料の解析－. 立命館大学地理学教室編『観光の地理学』文理閣, 84-110.
- 矢部直人・有馬貴之・岡村 祐・角野貴信 (2010). GPSを用いた観光行動調

位置情報付き Twitter データの観光行動分析への有用性と限界 (桐村)

査の課題と分析手法の検討. 観光科学研究, 3, 17-30.

渡辺隼矢 (2016). 位置情報付き Twitter 投稿データを利用した観光行動分析の手法開発. 地理情報システム学会講演論文集, 25, CD-ROM.

Goodchild, M. F. (2007). Citizens as sensors: the word of volunteered geography. *GeoJournal*, 69, 211-221.

Morstatter, F., Pfeffer, J., Liu H. and Carley, K. M. (2013). Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, 400-408.

注

- 1) <https://twittercommunity.com/t/foursquare-location-data-in-the-api/36065> (最終閲覧日: 2017年9月1日)
- 2) <https://www.biglobe.co.jp/pressroom/release/2014/06/140609-a> (最終閲覧日: 2017年9月1日)
- 3) 特定の店舗や場所を訪問したことを投稿し、その場所に関する特典を得たり、他のユーザーとその体験を共有したりできるサービスを指す。
- 4) 前掲1)。
- 5) 三重県が定める伊勢志摩定住自立圏の範囲であり、伊勢市、鳥羽市、志摩市、玉城町、度会町、大紀町、南伊勢町、明和町からなる。
- 6) 第1表に示したもののほか、Windows Phone やその他のタブレット用の Twitter アプリ、Instagram のアプリなどを対象とした。
- 7) <http://www.stat.go.jp/data/kokusei/2015/users-g/pdf/04.pdf> (最終閲覧日: 2017年9月13日)
- 8) <http://www.csis.u-tokyo.ac.jp/UEA/> (最終閲覧日: 2017年9月13日)
- 9) <http://www.city.ise.mie.jp/secure/34591/H27kekagaiyoukihon.pdf> (最終閲覧日: 2017年9月13日)
- 10) 居住地判定では、日本国外での投稿を考慮しておらず、このユーザーには外国人観光客も含まれている。

Usefulness and Limitation of Geotagged Twitter Data in Tourist Behavior
Analysis : A Case Study of the Ise-Shima Region

Takashi KIRIMURA

Abstract

This study aims to reveal the usefulness and limitation of geotagged Twitter data, which is representative of geotagged social networking services (SNSs) log data, in tourist behavior analysis. Twitter users' tourist behavior can be traced by using their ID numbers. This research utilizes data from geotagged tweets posted in Japan through mobile Twitter apps between April 2012 and March 2015, when coordinate-based geotagged Twitter data accounted for most of the available data. There are two major issues as follows.

First, geotagged Twitter data were used in various types of geographic research. However, the geographic characteristics of the data have not been sufficiently discussed in previous research. The residences of Twitter users in Japan, determined from their past geotagged tweets, tended to be concentrated in the Tokyo and Osaka metropolitan areas. In a case study of Ise, one of the major tourist cities in Japan, Twitter users who visited the city lived in these metropolitan areas more intensively than the residences of actual tourists known by the tourist survey by the Ise city office. If we use geotagged Twitter data for the purpose of tourist behavior analysis, then either the analysis needs to be based on the Twitter users' residences or the data need to be modified.

Second, Twitter Inc. had changed the type of location information associated with geotagged tweets from "coordinates" to "places" in late April 2015. Post implementation, the coordinates associated with the geotagged

位置情報付き Twitter データの観光行動分析への有用性と限界 (桐村)

tweets had become originated from check-in web services such as Foursquare. The spatial units that we could use for analysis also changed into places that mostly corresponded to municipalities in Japan. As a result of the analysis of tourist behavior in the Ise-Shima region, both the inter-municipal tourist flow, such as the one between the Ise city and the Toba city, and the tourist flow within a single municipality, such as the one between Naiku and Geku, were observed by analyzing coordinate-based Twitter data. The inter-municipal flow was considered to be analyzable using the place-based geotagged Twitter data that was available after late April 2015; however, the flow within a single municipality had not been sufficiently analyzed using place-based data. Currently, coordinate-based geotagged Twitter data have become a part of the entire geotagged Twitter data. Therefore, we need to analyze complete trends in tourist behavior using place-based geotagged Twitter data to precisely examine the inner flow within a single municipality.

Keywords : big data, social networking service (SNS), residences of Twitter users, three major metropolitan areas, tourist flow