

# 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—

桐村 喬

## Basic Characteristics of User Behavior in Geotagged Twitter Data:

### Applicability to Tourist Behavior Analysis

Takashi KIRIMURA

**Abstract:** This paper aims to examine the applicability of geotagged Twitter data to tourist behavior analysis. Twitter users in Japan post more geotagged tweets on Saturday and Sunday than other days. Each extent of the user's behavior is also wider on weekends than weekdays. We are able to calculate the number of visitors based on their daily living area. The daily living area indicates the area where they most posted in the past sixteen months. The data makes it possible to analyze spatiotemporal patterns of the visitors. The data, which is accessible by everyone, is useful for the tourist behavior analysis.

**Keywords:** 観光行動 (tourist behavior), 生活圏 (daily living area), 京都市 (Kyoto City)

#### 1. はじめに

140文字までの短文をウェブに投稿し、情報を発信・共有できるサービスであるTwitter（ツイッター）では、GPSによって求められた位置情報や場所の名称に基づく位置情報を、投稿に付与することができる。位置情報が付与される場面は、通勤途上や仕事場、旅先など様々であり、位置情報とともにユーザーの発言内容が投稿されている。ユーザーの日常の一端を垣間見ることのできるこのビッグデータは、日常的な行動や旅先での行動などを把握できるものであり、観光行動の分析にも活用できると考えられる。

観光行動の分析は、既存の観光統計のみからで

は把握しづらく、アンケート票による調査が主に行なわれてきた。近年は、GPS（杉本，2012）やIC乗車券（矢部・倉田，2013）などを利用した観光行動データの収集も行なわれている。ツイッターの投稿データからは、各ユーザーの居住地、あるいは日常的な生活圏の範囲を特定する必要があるものの、旅先などでの観光行動を、各ユーザーが投稿した範囲内で把握できると考えられる。また、投稿データは、ツイッターのユーザーであれば誰でも無料で入手可能であり、日本全国を対象としたユーザーの行動の分析も可能である。

そこで本研究では、位置情報が付与されたツイッターの投稿データを利用して、ユーザーの行動の基本的な特徴を整理し、観光行動分析への利活用の可能性を検討する。多数の研究者がアクセス可能であり、日本だけでなく全世界もカバーしたデータによる観光行動分析の実現可能性を示すことができれば、より多くの地域や観光対象に関

---

桐村 喬 〒113-8656 東京都文京区本郷 7-3-1

日本学術振興会特別研究員

(東京大学大学院工学系研究科 都市工学専攻)

Phone: 03-5841-6227

E-mail: kirimura@csis.u-tokyo.ac.jp

する観光行動の実態を解明できるようになる。

## 2. 投稿データの概要

ツイッターのサービスを提供する米国 Twitter 社は、投稿されたデータの一部をユーザーに公開している。一般のユーザーがアクセスできるデータはサンプリングされたものではあるが、位置情報が付与された投稿データに限れば、そのほとんどを得ることができる (Morstatter et al., 2013)。

本研究では、筆者が 2012 年 2 月以降収集してきた日本とその周辺の位置情報が付与された投稿データのうち、2013 年 5 月末までの投稿データを利用する。収集を行なうプログラムの不具合などのために若干の漏れが生じたものの、1 年 4 ヶ月間に収集した投稿データは約 5,600 万件であり、このうち日本国内の陸上の位置情報が付与されたものは約 4,500 万件である。

表-1 大都市圏別にみた投稿データの位置情報

| 地域    | 投稿数<br>(2012年2月-<br>2013年5月) | 人口<br>(2010年<br>国勢調査) | 1,000人<br>あたり<br>投稿数 |       |
|-------|------------------------------|-----------------------|----------------------|-------|
| 三大都市圏 | 東京                           | 16,603,053            | 36,534,274           | 454.5 |
|       | 名古屋                          | 2,763,866             | 9,107,414            | 303.5 |
|       | 京阪神                          | 8,034,126             | 19,341,976           | 415.4 |
| その他   | 17,990,158                   | 63,073,688            | 285.2                |       |
| 合計    | 45,391,203                   | 128,057,352           | 354.5                |       |

投稿データに付与された位置情報は、人口の集中する大都市圏のものが中心であり、とりわけ二大都市圏に偏っている (表-1)。しかし、市区町村別に 2010 年の国勢調査による昼間人口 1 人あたりの投稿数を求めると、その上位には、岐阜県白川村 (白川郷) や岩手県平泉町 (中尊寺)、新潟県湯沢町 (スキー場、温泉) など、著名な観光地や観光資源を抱える市区町村が散見される。これらの地域では、居住者や従業者・通学者だけでなく、観光客を中心としたユーザーによる投稿も多いと推測され、多くのユーザーにとって観光は、位置情報を付与してツイッターに投稿する主要

な場面の 1 つであると考えられる。

## 3. ユーザー単位でみた行動範囲

### 3.1 ユーザーの生活圏の判定

観光行動の分析を行なうためには、各ユーザーの投稿データのなかから、ユーザーの日常的な生活圏を判定し、日常的な生活圏ではない場所での投稿を抽出する必要がある。どの程度の地理的な単位を日常的な行動範囲と考えるかは分析目的によって異なるが、ここでは簡単に都道府県を基本単位として、最も投稿数の多い都道府県をユーザーの生活圏と判断する。なお、計算を容易にするために各投稿データの位置情報が示す都道府県の判定は 1km メッシュ単位で行なった。投稿データに含まれる約 110 万ユーザーのうち、生活圏が日本国内と判断されたのは約 87 万ユーザーである。都道府県別にみれば、二大都市圏に属する都道府県を生活圏とするユーザーが多く、昼間人口あたりでは東京都と京都府が多い (表-2)。

表-2 生活圏別にみたユーザー数上位 10 地域

| 都道府県 | ユーザー数<br>(2012年2月-<br>2013年5月) | 昼間人口<br>(2010年国<br>勢調査) | 昼間人口<br>1,000人<br>あたり<br>ユーザー数 |
|------|--------------------------------|-------------------------|--------------------------------|
| 東京都  | 192,580                        | 15,576,130              | 12.4                           |
| 神奈川県 | 72,991                         | 8,254,193               | 8.8                            |
| 大阪府  | 71,625                         | 9,280,560               | 7.7                            |
| 愛知県  | 46,521                         | 7,520,876               | 6.2                            |
| 千葉県  | 45,742                         | 5,560,489               | 8.2                            |
| 埼玉県  | 43,020                         | 6,373,489               | 6.7                            |
| 兵庫県  | 33,028                         | 5,347,839               | 6.2                            |
| 北海道  | 32,100                         | 5,504,418               | 5.8                            |
| 福岡県  | 27,202                         | 5,078,054               | 5.4                            |
| 京都府  | 24,999                         | 2,668,371               | 9.4                            |

### 3.2 ユーザー行動の曜日差

観光のために日常的な生活圏とは違う場所を訪れるのは、多くの場合、週末や長期休暇の期間である。そこで、ユーザーの行動範囲の特徴を把握するために、曜日ごとにユーザーの移動距離の

平均値を求め、その値の中央値を算出した(図-1)。すべての曜日で移動距離を求めることができるユーザーは約11万ユーザーであり、生活圏を日本国内とするのは約9万ユーザーである。日本国内のユーザーの移動距離は、週末で長く、週の中盤で短くなっており、週末ほど行動範囲が広いことがわかる。生活圏別にみると、最長の曜日と最短の曜日の差は、島根県や宮崎県など、二大都市圏以外の都道府県ほど大きい傾向がみられた。

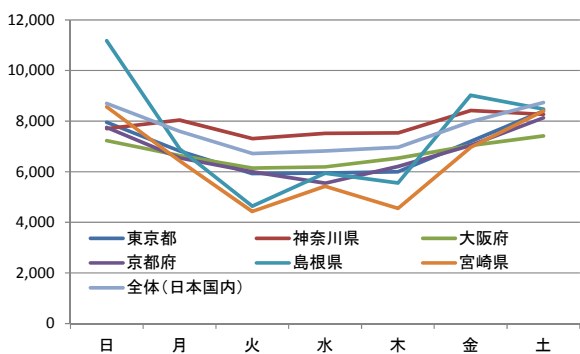


図-1 曜日別の移動距離の中央値 (単位: m)

## 4. 京都市におけるユーザー行動の分析例

### 4.1 京都市への訪問ユーザーの抽出

これまで整理してきたように、ツイッターユーザーの行動範囲は、週末により広がる傾向にあり、観光や買い物など余暇に伴う各ユーザーの行動の一端が投稿データに反映されている可能性が高い。そこで、日本を代表する観光都市である京都市に注目して、観光行動に関する若干の分析を行ない、位置情報付きのツイッター投稿データの利用可能性や問題点について整理、検討する。

まず、投稿データから京都市への訪問者であるユーザーを抽出する。ここでは、京都市内で1度でも投稿したことのあるユーザーのうち、京都市内を生活圏とせず、かつ京都市内での投稿数が全投稿数の1割以下であるユーザーを、京都市への訪問ユーザーとする。

訪問ユーザー数は22,389であり、地方単位での生活圏別にみれば近畿地方(37.9%)と関東地

方(36.8%)が多い。2010年の京都市観光調査の結果に基づく観光客数の出発地別の割合は、近畿地方が最も多く(61.2%)、関東地方(14.0%)、中部地方(11.5%)と続いており、訪問ユーザーは観光客の分布よりも関東地方に偏っている。

### 4.2 京都市への訪問ユーザーの訪問地の特徴

訪問ユーザーの京都市内での訪問地を1kmメッシュ単位でみれば、京都駅を含むメッシュに全体の42.0%のユーザーが訪問しており、それに次ぐのは繁華街の四条河原町を含むメッシュ(19.0%)である。一方、京都市観光調査によると、最も観光客の多い清水寺(21.0%)に対して、京都駅(4.6%、ただし京都駅ビルとして)や四条河原町(5.4%)への訪問はそれほど多くない。

投稿データには投稿時間の情報が含まれており、任意の時点・地域における訪問ユーザー数を把握できる。訪問ユーザー数が最も多い京都駅を含むメッシュは、12時台と18時台にピークがあるものの、6時台でも多くなっており、交通の結節点としての特徴が表われている(図-2)。四条河原町を含むメッシュなど、繁華街を含むメッシュでの訪問ユーザー数は、13時台から18時台前後をピークとし、夜間も比較的多い。一方、観光客の主要な訪問地である社寺を含むメッシュでは、拝観時間の関係から、9時台から12時台にかけて急速に増加し、16時台以降で急速に減少するパターンを示している。

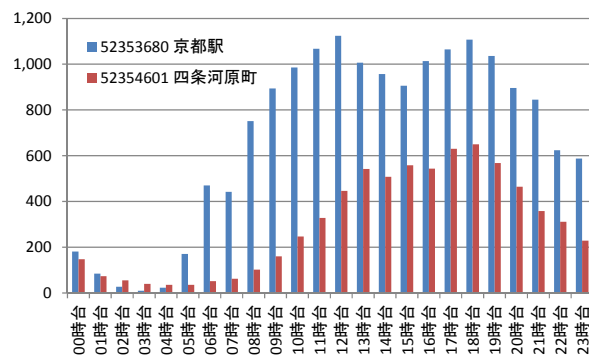


図-2 時間帯別の訪問ユーザー数 (京都駅および四条河原町を含む各メッシュ)

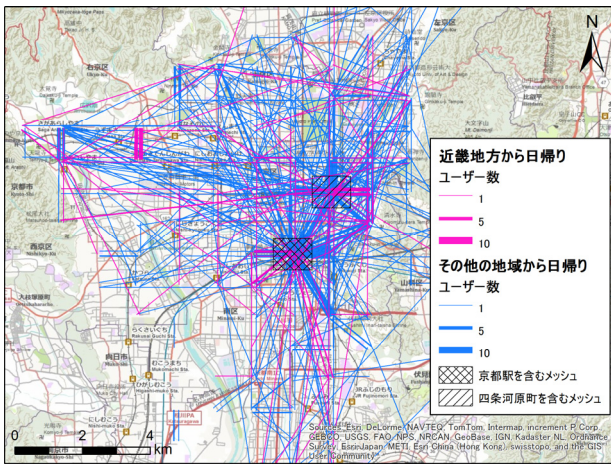


図-3 京都市への日帰り訪問ユーザーの流動

#### 4.3 京都市への訪問ユーザーの市内移動

次に、訪問ユーザーの京都市内での移動をメッシュ単位で視覚化する。対象とする訪問ユーザーは、データの収集期間中に、京都市内での投稿が1日のみであり、10回以上投稿した日帰りの180ユーザーである。生活圏が近畿地方である43ユーザーは、京都駅を含むメッシュだけでなく、四条河原町を含むメッシュも移動の拠点となっているものの、その間の移動は少ない(図-3)。一方、残り137ユーザーについては、京都駅を含むメッシュと四条河原町を含むメッシュとの間の移動量が多く、また、京都駅を含むメッシュを起終点とした主要観光地への移動が多い点が、近畿地方を生活圏とする訪問ユーザーと異なる。

#### 5. おわりに

本研究では、位置情報が付与されたツイッターの投稿データを利用して、ユーザーの行動の基本的な特徴を把握し、京都市を事例とした観光行動の分析事例を示した。

ユーザーの日常的な生活圏は二大都市圏にやや偏っているものの、その行動範囲は週末により広がる傾向が確認され、観光などの余暇活動の一端がツイッターの投稿データに表われていると考えられた。京都市を事例とした観光行動に関する若干の分析からは、生活圏別の訪問ユーザー

数や、市内の主要駅や観光地における時間帯別の訪問ユーザー数を求めることができた。また、京都市内でのユーザーの移動の状況も把握でき、近畿地方からの日帰り客と、それ以外の地域からの日帰り客の行動の差異を示すことができた。

このように、位置情報が付与されたツイッターの投稿データからは、一定の観光行動を把握することができることが示された。しかし、投稿データ自体からの生活圏や交通手段などの推定は可能であるものの、ユーザーに関する詳細な属性を入手することは難しく、年齢や性別、同行者数などによる行動の違いにまで言及することは不可能である。この点は大きな問題ではあるが、全国あるいは世界的に同様のデータが容易に入手可能であり、広範囲を対象とした観光行動の分析が可能であることは魅力的である。例えば、ユーザー単位で年間の行動を追跡すれば、観光行動に関する個々のユーザーの嗜好も把握できる。

今後は、他の統計資料などを活用した投稿データに基づく属性の推定方法の開発や、プログラムなどによる機械的な投稿データを排除する方法の検討など、より分析の精度を高めるための研究を進めていきたい。

#### 謝辞

2012年2月から2013年2月までのデータ収集には、立命館大学地理学教室のサーバを利用させていただいた。ここに記して厚くお礼申し上げます。

#### 参考文献

- 杉本興運 (2012) : 観光者の視覚的体験情報に基づく回遊空間の評価, GIS-理論と応用, 20(1), 39-50.
- 矢部直人・倉田陽平 (2013) : 東京大都市圏におけるIC乗車券を用いた訪日外国人の観光行動分析, GIS-理論と応用, 21(1), 35-46.
- 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.