

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

鈴木英之

Studies on Destination Rating by Location Data on Social Media

Hideyuki SUZUKI

Abstract: The main purpose of this paper is to the study of the use of the data of social media for the rating of tourist destinations. For this purpose, the data derived from the social media, it was necessary to investigate the basic characteristics and the bias. Result of investigation, Posts of social media was found to be a strong correlation with the rating of tourist destination. In addition, it was found to have undergone a certain effect on the non tourist factors, as well as tourism factors.

Keywords: ソーシャルメディア (social media), フリッカー (Flickr), ツイッター (Twitter), 観光資源 (tourism resources), 観光地ブランド (destination brand)

1. はじめに

1.1 背景

位置情報付きソーシャルメディアや移動体データから類推される観光行動の分析レポートングを目的とするサービスが、地域経済分析システム RESAS やその他民間事業者によって実用化され始めている。しかし、それらが依拠するデータは統計的な調査法を通して得られたものではないため、その代表性に問題が無いとはいえない。また、現実の観光行動の実態を適切に反映したのかどうかという点についても今後の検証が待たれる。

1.2 目的

本報告は、位置情報付きソーシャルメディア等のデータを用い、観光行動の分析を実施する際に、それらがどのような固有の特徴や偏りを有するのかという点を明らかにすることで、観光地の格付け等の評価を行う際の留意点として検討することを目的とする。

2. 方法

まず、観光地の「格」の概念について整理し、続いて使用データと方法の概略について説明する。

2.1 観光地の格付け

観光地の格付け（魅力度・ブランド）評価については、これまで、研究者や実務家によって様々な方法が提案されてきた。多くに共通するのは、還元主義的な手法である。まず観光資源の種別や多角的な観点に沿って細分化された評価項目のそれぞれに対して、個別評点を下し、次いでそれらを積み上げることで総合評価とする方法だ。最終的な総合評価は抽象度が高く、観念的に理解される場合も少なくないものと思われるが、入込数や集客圏といった現実の変数との対応関係を想定するものもある。田村（2012）は小売吸引力モデルを援用し、それぞれの観光地をその集客範囲によって、3階層構造の中に位置づけることで分類した。また、公益財団法人日本交通公社は、実証的な分析に基づくものではないにせよ、集客圏に応じた観光資源の階層的分类を行っている（後述）。

本報告では、観光地の「格」をそれぞれの観光地の集客圏の広さで定義する。操作的には外国人観光客の来訪者の規模と同義とみなし、ソーシャルメディア等における外国人投稿数の観測値で、十分に実用的な測定が可能であろうと仮定する。

2.2 使用するデータ

2.2.1 ソーシャルメディア等データ

ソーシャルメディア等のデータとして、下記2種類の座標付きデータを利用する。多頻度投稿者によるデータの歪みを防ぐために、同一投稿者の投稿を一定時間内の投稿頻度の逆数で重み付けしたものを集計して用いる。

・ Flickr

2012年1月から2016年6月の間で国内に位置情報を持つ投稿を対象とする。投稿者の居住地が判明している230万件（内、外国人によるもの80万件、日本人によるもの150万件）及び、居住地が判明しない220万件的合計450万件を利用する。

・ Twitter

2016年6月20日から8月20日の間で、国内の座標を付されて投稿されたものの内 Streaming API から取得したものを利用する。中谷(2015)を参考とし、昼間の時間帯（8:00-18:00）のものを利用する。使用言語により日本語使用者（240万件）及び外国語（日本語以外の言語）使用者（35万件）に分類し利用する。

2.2.2 既存の格付けデータ

観光地の格付けに対して、既存の下記2種類の資料を参照する。

・ 観光資源台帳

観光資源台帳は、(財)日本交通公社が、全国の個別の観光資源に対して、格付け（等級）評価を付したリストである。1977年に作成された後、1999年（以降、旧版）、2014年（新版）に改訂版が発表されている。世界レベルの観光地点を意味する特A級から、近隣住民の観光対象であるC級レベルまで、階層的な観光地体系を仮定した内容となっており、上位階層（旧版では特A級、A

級、B級の混在する2,019件、新版では特A級、A級に分けて計451件）については、一般公開されているものの下位の観光資源については諸般の事情により公表されていない。

そこで、旧版リストから新版の特A、A級を除外したものを、B級資源とみなす、という変則的な方法によって3階層別の観光資源リストを用意した。

台帳に掲載される観光資源カテゴリーのうち、山岳、河川、景観、動物、植物のように対象となる空間領域が明瞭ではないもの、年中行事やイベント、食など空間的な集計に馴染まないものは除外し、特A級47件、A級112件、B級582件の観光地点を分析対象とする。

・ トリップアドバイザー

トリップアドバイザーは、米TripAdvisor社によって運営される旅行に関する世界的な口コミ評判サイトである。明示的に観光地の格付けを目的とした情報提供を行っている訳ではないが、投稿数や投稿内容を加味した独自の計算法による観光地ランキングを公表している。ただし、このランキングの趣旨は人気度（投稿件数）の他に満足度（口コミ内容及び評点）を合成したものとなっており、本報告が想定する観光地の「格」とは異なる。そこで、日本語以外の言語による投稿件数を観光地の「格」に対応する観測変数とみなした。国内33,000以上に及ぶ観光地の中から、総口コミ件数が50件以上のもの3214件（2016年6月現在）の観光地点を分析対象とする。

2.3 方法の概略

まず、「既存の格付けデータ」の観光地各地点に対するソーシャルメディア等データの投稿件数を集計し、観光地の「格」との対応関係について確認する。空間的な集計範囲は既存の観光地の座標から半径500mの円バッファとする。

当然ながら、ソーシャルメディア上の行動（投稿）は、観光行動のみに伴って発生する訳ではない。観光以外の社会経済的な行為（地域在住者の

日常的なルーティン行動のみならず、観光客の非観光行動等)を反映した投稿が一定割合で混在しているものと考えられる。

そこで、空間における観光的要因と非観光的要因のそれぞれがソーシャルメディア上の投稿件数に、どの程度の影響を与えるのか、回帰分析によって確認する。具体的には、2分の1地域メッシュを集計単位としてソーシャルメディア投稿数に対して、観光行動の観測変数としてトリップアドバイザー口コミ件数及び、非観光的な社会経済活動の代理変数としての平成24年経済センサスの就業者数を説明変数とする回帰モデルに対してパラメータの比較を行う。

3. 結果と考察

3.1 観光資源台帳の観光地における集計

観光地点別に集計した結果を等級別に比較する。Flickr, Twitter とともに観光地の格(等級)が高くなるに依りて、投稿数は増加する傾向がみられる。この傾向は、投稿者の外国人、日本人の別を問わず共通している。

中でも、外国人による Flickr の投稿数は、下位等級(B級)に対する上位等級(特A級)の比で、178倍にも及び、観光地の格を評価する際の識別力が強いと言える。

ただし、等級が下がるにつれて、投稿件数の減少が目立つ。B級では、外国人の投稿数がゼロである観光資源が Flickr, Twitter 共に4割以上に及び、分析上の支障となる。日本人による投稿数を利用した外国人数の推定や、複数のソーシャルメディアのデータを合成する等の工夫が求められるものと思われる。

表1 観光資源台帳等級別ソーシャルメディア投稿件数の中位数(上段)と投稿数ゼロの観光資源比率(下段)

	Flickr			Twitter : 昼間			参考) TripAdvisor		
	外国人	日本人	全数	外国語	日本語	全数	外国語	日本語	全数
特A級	177.6	76.3	428.1	32.0	197.6	259.8	115.0	1096.0	1316.0
	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
A級	60.1	70.5	245.7	15.3	116.9	129.9	129.0	438.5	575.5
	6.3%	1.8%	0.9%	11.6%	2.7%	1.8%	10.7%	4.5%	4.5%
B級	1.0	8.0	20.4	1.0	15.0	16.3	3.0	45.5	49.0
	41.4%	14.6%	7.2%	45.7%	8.6%	8.6%	33.7%	14.4%	14.4%

3.2 TripAdvisorの観光地における集計

ここでは、TripAdvisorに投稿された外国人口コミ件数を十分位でクラス化し、階級別のソーシャルメディア投稿件数の分布を観察する。

前項同様に、上位の観光地ほど投稿件数の増加が明瞭に確認される。やはり外国人 Flickr 投稿数において、第1十分位等級に対する第10十分位等級の比は600倍を超え、優れた識別力を示した。

表2 トリップアドバイザー口コミ十分位等級別ソーシャルメディア投稿件数の中位数

十分位	TripAdvisor			Flickr			Twitter : 昼間		
	外国語	日本語	全数	外国人	日本人	全数	外国語	日本語	全数
10	291	352	678	316.5	276.1	1245.4	118.0	612.4	729.3
9	87	152	250	86.0	119.0	404.8	31.9	228.2	269.2
8	43	122	170	45.1	77.8	258.6	22.4	173.6	202.3
7	26	108	136	23.7	37.6	129.4	11.1	108.2	119.2
6	17	84.5	102	14.3	36.9	109.3	6.9	68.1	81.4
5	11	78	90	9.0	26.3	78.3	8.0	78.4	93.0
4	8	72.5	80	4.9	17.1	44.0	3.2	42.7	45.8
3	5	67	72	2.0	12.4	30.9	2.6	35.5	39.1
2	2	63	66	1.0	8.5	21.2	1.5	26.5	28.0
1	1	62	63	0.5	6.4	15.9	1.0	25.3	26.5

3.2 非観光要因の効果

これまで見てきたように、ソーシャルメディアの投稿件数と、観光行動のボリュームは、概ね比例関係を有するよう見える。しかし双方に影響を与える潜伏変数の確認を行う必要がある。ソーシャルメディアの投稿件数が、都市部に偏在することはよく知られており、観光地の所在地もまた、同様の空間的な偏りがあるとすれば、そのことがもたらす影響も無視できないからである。

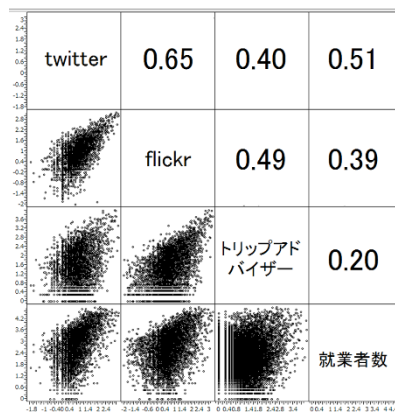


図1 2分の1地域メッシュで集計した各変数の相関行列

都市中心部における社会経済活動を表現する代理変数として、経済センサス就業者数を用い、主として都市部における非観光要因を説明する

変数とみなして、ソーシャルメディア投稿件数や観光行動に関する変数との関係を調べた。

相関係数で見ると、観光要因(トリップアドバイザー口コミ件数)と非観光要因(経済センサ

ス就業者数)の相関は、さして高くない(図1)。
 潜伏変数の問題は考えなくても良さそうである。

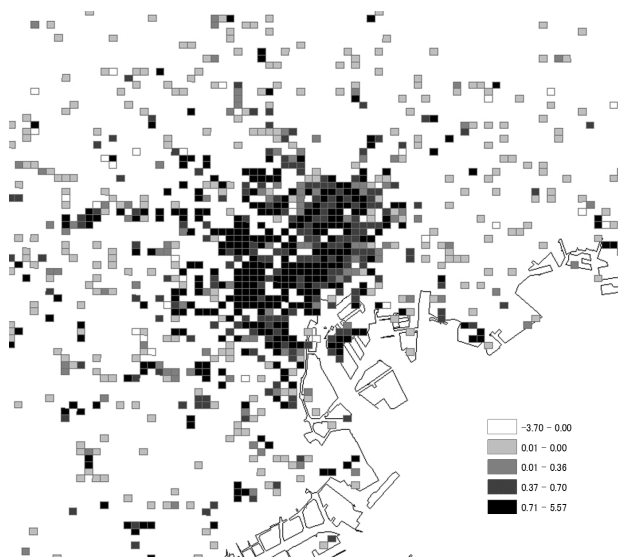


図2 外国人によるトリップアドバイザー口コミ件数に対する外国人による Twitter 投稿数の比

またソーシャルメディアの投稿件数は、Flickr, Twitter とともに観光要因、非観光要因の双方と相関が見られる。特に都市部においては、非観光要因が一定のウェイトを占めることが推察される。

図2は、ソーシャルメディア投稿件数を観光口コミ件数で基準化した値を示したものである。都市中心部で大きな値を示す。この傾向は、三大都市圏及び主要な地方都市において認められる。この値の大きさは、ソーシャルメディア件数に対する観光要因以外の要因を示唆するものと解釈することが可能である。

表3 ソーシャルメディア投稿数に対する回帰分析結果

	Twitter					Flickr				
	B	SE B	β	t値	p値	B	SE B	β	t値	p値
切片	-0.57	0.03		-20.28	<.0001	-0.52	0.03		-17.03	<.0001
就業者数	0.34	0.01	0.47	34.86	<.0001	0.29	0.01	0.35	26.68	<.0001
トリップアドバイザー	0.34	0.01	0.40	29.52	<.0001	0.56	0.01	0.55	42.49	<.0001
R ²	0.45					0.48				

最後に、ソーシャルメディアの投稿件数を観光要因(トリップアドバイザー口コミ数)と非観光要因(就業者数)の2要因で説明する回帰分析を

適用し、パラメータの比較を行う。2つのソーシャルメディアはそれぞれ、2要因のそれぞれから影響を受けている。標準偏回帰係数を比較すると、Twitter では観光要因より非観光要因の効果が大きく、Flickr では観光要因の効果のほうが大きい。

いずれにせよ、ソーシャルメディアの投稿数で観光地の格付けを行う際には、就業者数等の非観光要因を示す変数で層別した分析を行うか、統制変数に加えて調整する等の考慮が必要であろう。

5. おわりに

ソーシャルメディアから得られるデータのようにノイズの混入や無視できないバイアスを抱える場合の対処としては、データのクリーニングや、目的に応じたフィルタリングを入念に行うのが常道であろう。ただし、場合によっては多くのサンプルを犠牲にしなければならず、結果として小規模な観光地や、よりマイクロな空間単位を対象とする際に度々困難が生じる。本報告では、獲得されたデータをそのまま利用しなくてはならない場合に、分析上の支障となるバイアスを事前に理解することを念頭に置いている。同様の状況に置かれた分析者の参考となれば望外である。

参考文献

- 大久保航太郎ら(2015): 複数の位置情報付きソーシャルメディアから得られたPOIの比較, 情報科学技術フォーラム講演論文集, 14(4), 481-482.
- 田村正紀(2012): アメニティが決める観光地の魅力, 「観光地のアメニティ」, 白桜書房, 1-48.
- 大井達雄(2013): 観光地ブランドの評価に関する一考察, 日本統計研究所報 (42), 9-27.
- 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, 「観光の地理学」, 文理閣, 84-110.
- 溝尾良隆(2014): 観光資源, 観光地の魅力評価の系譜: 誰が評価してきたのか, 観光文化, 38(3), 3-9.
- 倉田陽平ら(2010): あえて案内しない着地型観光案内-観光関心点群の抽出と活用, 地理情報システム学会講演論文集, 19, CD-ROM.