

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

渡辺 隼矢

Development of Analytical Method of Tourists' Behavior Using Geotagged Twitter Data WATANABE Junya

Abstract: Twitter data with Geo-tag is useful for the analysis of tourists' behavior through classifying users based on the length of stay in a city. This paper tried to propose two analytical methods about tourists' behavior in Kanazawa-City, which are hot-spot extraction and visualization of tourists' movement. The results show the centrality of Kanazawa-station. In the future, estimation of users' attribute and combination with other method such as questionnaire survey are expected.

Keywords: 都市観光 (urban tourism), 観光行動(tourists' behavior), SNS, Twitter, 金沢市(Kanazawa-City)

1. はじめに

近年ソーシャル・ネットワーク・サービス(以下 SNS)の投稿データを利用して観光に関する分析を試みた研究が観光地理学, 都市工学, 情報工学分野等で見られるようになった。これらの特徴として, 不特定多数の人々から発信された大量の投稿データを収集して分析を行っている点, 投稿に付与された位置情報データを利用して, ミクロスケールな場所における分析を行っている点等が挙げられ, 今後もこれらの研究は増加すると考えられる。中でも, Twitter の投稿データは特に着目されており, 事例を挙げると, 桐村(2013)は, 投稿における位置情報の移動から, 京都市における日帰り観光客の分析を行い, また観光庁(2015)においても, 日本の地名や旅に関する語を含む英語の投稿の抽出から, 訪日外国人観光客の興味・関心の内容の把握を試みた。

しかしながら, これらの先行研究は Twitter 投稿データの活用事例の範に留まっており, 分析手法の検討はなされていない。今後の観光行動分析への普及に向けては, 手法や適用例を含めた研究の蓄積が不可欠である。そこで本研究は, 2015年3月の北陸新幹線開業以降, 観光ブームが発生している金沢市を対象に, Twitter の位置情報付き投稿から短期滞在者の行動を可視化するとともに, 同市における観光動態の時空間構造を分析する手法を開発することを目的とする。また分析を通して, これらの投稿データを利用した観光行動分析の今後の可能性や課題についても考察したい。

2. Twitter 投稿データについて

2.1 データの取得・抽出

Twitter の投稿データは, サービス運営主体である Twitter 社が提供する API を利用することで, 誰でも簡単に投稿データを収集することができる(桐村, 2013)。またその際, ユーザー名や投稿場所, 言語等を指定して収集することが可能である。

渡辺隼矢 〒305-8572 茨城県つくば市天王台 1-1-1

筑波大学大学院 生命環境科学研究科

E-mail: s1620976@u.tsukuba.ac.jp

本研究では2015年5月から同年10月までの半年間、金沢市での位置情報付きの投稿を収集し、58,567件が集められた。

短期滞在者の抽出方法には、ユーザー毎の投稿日時に着目した佐伯ほか(2015)の手法を一部修正して活用した(図-1)。具体的には、まずはデータ収集期間をD日毎に区切り、D日間を1つのブロックとする。期間内投稿数がT以上のユーザーのうち、投稿が確認されたブロックの数がR以下のユーザーを短期滞在者として抽出した。またRより大きいユーザーのうち広告・bot等を除いたものを長期滞在者として識別した。本研究では閾値を、 $T=3$ 、 $D=3$ 、 $R=3$ と設定した。佐伯ほか(2015)においては、適切な閾値を設定することで、訪日・在日外国人を9割以上の精度で分類することができた。本研究では閾値の検証は省略したが、投稿内容等を見ると、短期滞在者として抽出されたユーザーの投稿は、概ね観光客の投稿と考えられるものであった。

2.2 分析対象データの概要

前節における分類の結果、553の短期滞在ユーザーが抽出され、それらのユーザーによる7,634件の投稿が本研究における分析対象となる。なお、長期滞在者として識別されたユーザー数は654で、投稿数は31,731件であった。

時間帯別の投稿数を見ると(図-2)、長期滞在者の投稿は夜間に多く、また食事時に投稿数が急増するのに対し、短期滞在者の投稿は日中に多く、夜間が少ないことが分かる。日別の投稿数では(図-3)、平日・土休日による若干の規則性はあるものの、観光客の多寡と必ずしも一致していないことが分かる。これはユーザー毎の市内における投稿数が不揃いであり、一人でより多くSNSへ投稿するような人の訪問数が結果に左右するためである。短期滞在者・長期滞在者ともに投稿数が最大であったのは10月10日で、この日は湯涌温泉が舞台のアニメ「花咲くいろは」のファンが多く現地に集ま

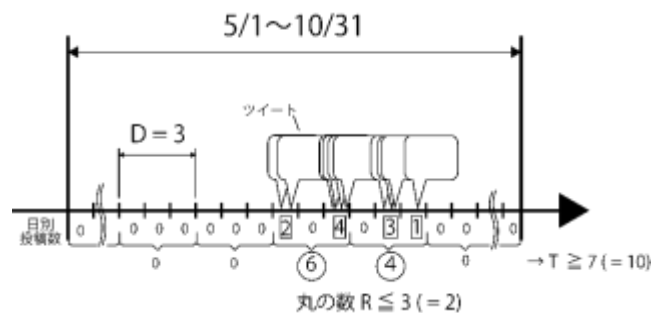


図-1 短期滞在者抽出モデル
(佐伯ほか(2015)を基に作成)

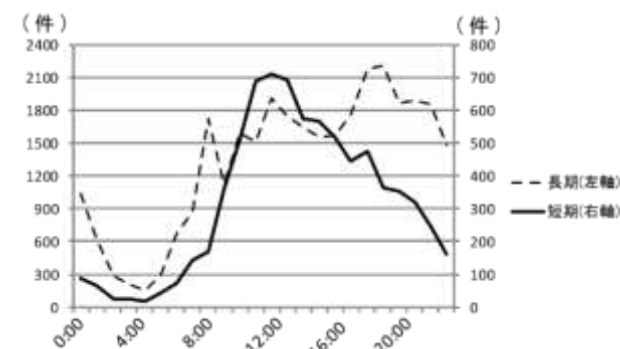


図-2 時間帯別投稿数

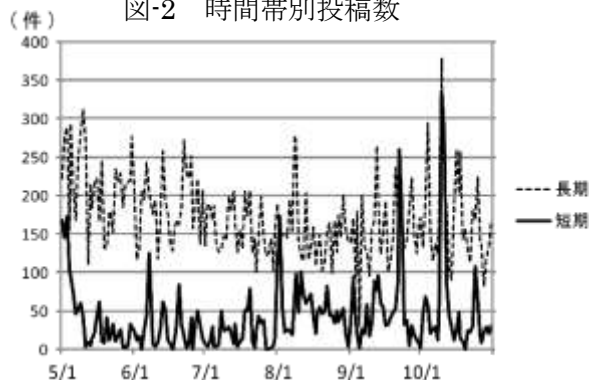


図-3 日別投稿数推移

る「湯涌ぼんぼり祭り」の開催日であった。

KHCoderを利用して投稿本文(写真共有サイト等の場所タグを除く)のテキストマイニングを実施した。その結果、名詞では「今日」(110回)や「今」(77回)、「到着」(45回)などが、動詞では「来る・くる」(203回)や「見る・観る・みる」(107回)、「食べる」(85回)などが多く登場したほか、「寿司」「カレー」「プール」「降る」など金沢を象徴する語も頻出語として登場した。

3. 分析と結果

3.1 投稿ホットスポットの抽出

各投稿に付与された位置情報から、金沢市内に

において短期滞在者の訪問および投稿地点の傾向を検証する。地理学ではポイントデータの集合を可視化する際に、カーネル密度推定を利用することが多いが、本研究では、ArcGIS の空間統計ツールである最適化ホットスポット分析を利用した。これを利用することで、指定したメッシュ毎に密度分布を計算し、そしてポイントが集中するホットスポットを抽出することができる。本研究では金沢市域に 100m メッシュを作成し、メッシュ毎にスコア値を算出した。

また分析に先立ち、同一ユーザーによる短時間での連続した投稿が分析結果に与える影響を小さくすることを目的に、データの離散化を行った。SNS の利用目的は人それぞれであるが、一人のユーザーが同じ場所で連続的に投稿をした場合、その地点の投稿密度は高くなるものの、短期滞在者が多く訪問する場所とは言い難い。このようなケースを排除するため、本分析では対象データに「任意の投稿 A から 30 分かつ 300m 以内で次の投稿 B があつた場合、投稿 B は最適化ホットスポット分析の対象外する」という制約を設けた。結果と

して、短期滞在者の全投稿のうち、離散化によって 3,356 件のデータが排除され、残りの 4,278 件のデータから分析を行った。

分析の結果(図-4)、金沢駅周辺において最も投稿が集中している事が明らかになった。スコア値は金沢駅周辺で約 40.0~45.4、以下は香林坊・広坂周辺(約 32.0~44.3)、片町周辺(約 16.0~33.0)、近江町市場、(約 16.0~18.0)、ひがし茶屋街(約 11.0)の順で、これらは全てホットスポットとして抽出された。郊外では湯涌温泉がホットスポットとして抽出された他、ホットスポットには満たないものの、観光名所ではないが短期滞在者の投稿が集中する場所が数地点見受けられた。

3.2 移動の可視化

次に、ユーザー毎の投稿地点の移動から、短期滞在者の移動の可視化を試みた。手法としては、まずは金沢市を「兼六園」「片町」「けやき通り」等 35 のゾーンに分割し、ArcGIS 上で短期滞在者の各投稿とゾーンを空間結合した。投稿間隔が 12 時間以内のものを対象に、ユーザー毎の投稿ゾー

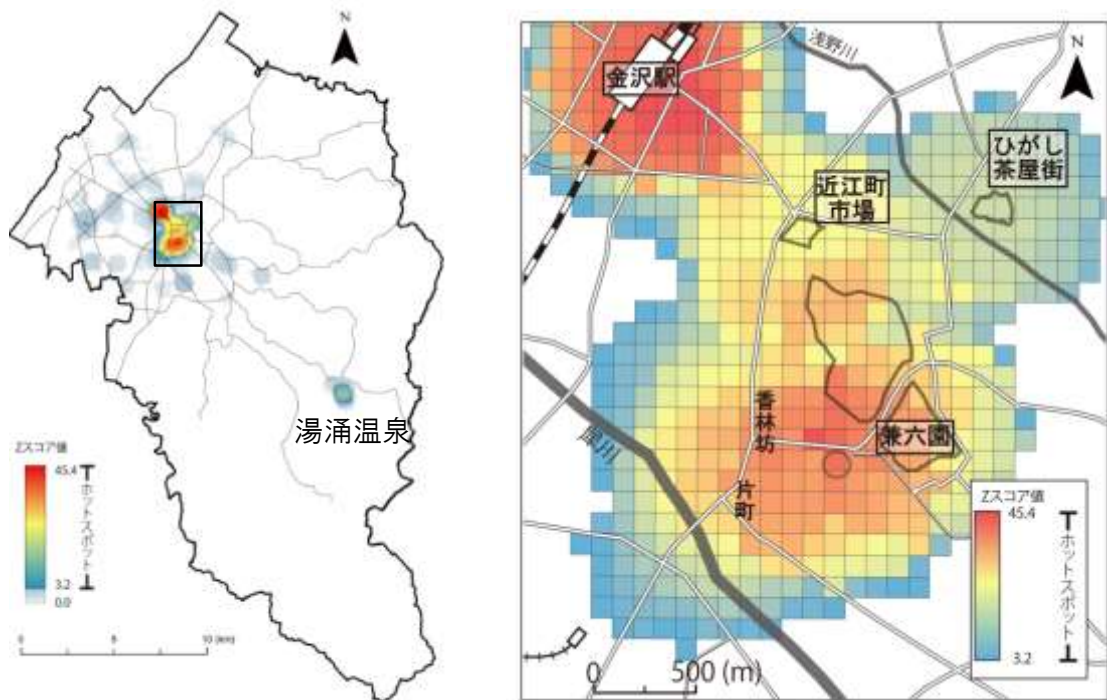


図-4 投稿ホットスポット分析結果

ン変化の OD 行列を作成した。

その結果(図-5),「金沢駅ゾーン」と他のゾーンを直接結ぶ移動が数多く見られ, 金沢駅の中心性が鮮明となった。トリップ数は「金沢駅ゾーン」から「近江町市場ゾーン」への移動が最大で, また「兼六園ゾーン」では隣接観光地への相互訪問が多い結果となった。ゾーン間移動を昼夜別で比べると(図-6), 日中は観光地間の相互移動が多いが, 夜間は観光地から金沢駅周辺や香林坊・片町などの繁華街へ向かう移動, そして駅および繁華街間の相互移動が多いことが見てとれた。

4. おわりに

本研究は金沢市において位置情報付き Twitter 投稿データから短期滞在者の投稿を抽出し, またそれを活用した観光行動分析の手法を 2 つ開発した。その結果, 短期滞在者のミクروسケールでの訪問先を明らかにすることができ, また移動状況も集計し, 可視化することができた。これらの位置情報を利用した分析結果は, 従来のアンケート方式で得るのが難しいものであった。

Twitter 投稿データを利用した観光行動分析の更なる普及のために, 今後はユーザー毎の投稿内容分析から, 年齢層や性別, 訪問目的といったユーザー属性の推定を進めるべきである。情報工学分野においてはこれらの研究が進んでおり, それらの積極的な活用が期待される。

また調査対象によって, SNS 投稿データが有用か否かは異なる。例えば, 兼六園においては, 訪問客の平均投稿数が少ないため, 園内における行動分析は難しいものの, 祭り訪問者を対象に, その前後の行動を分析するには SNS 投稿データは大いに役立つと考えられる。

今後は分析精度の検証や精度向上に加え, アンケート, GPS 端末など複数の媒体により収集されたデータを複合的に活用して観光行動を分析することが求められる。

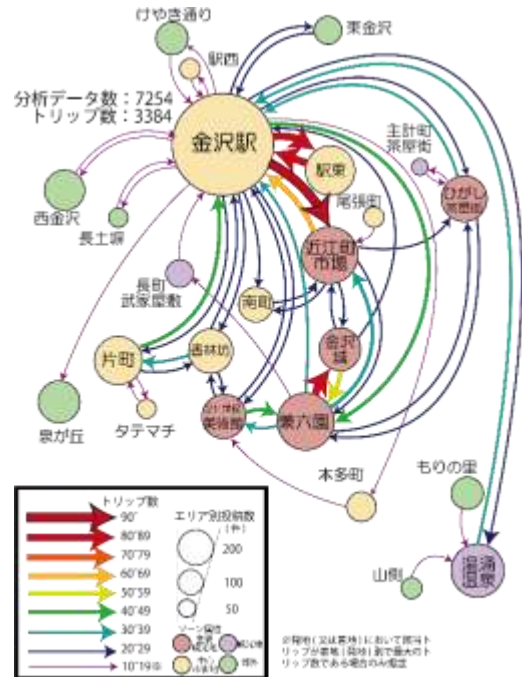


図-5 短期滞在者ゾーン間流動

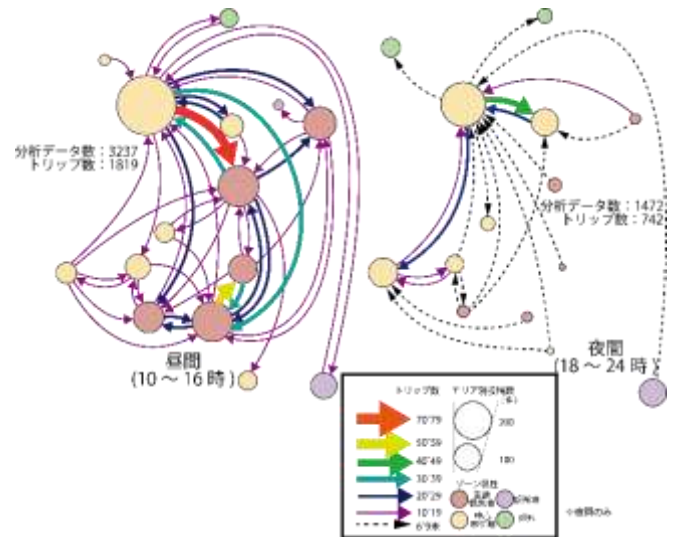


図-6 昼夜間別短期滞在者ゾーン間流動

参考文献

観光庁 (2015): ICTを活用した訪日外国人観光動態調査検討委員会資料。

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。

佐伯圭介・遠藤雅樹・廣田雅春・倉田陽平・横山昌平・石川博 (2015): 外国人 Twitter ユーザーの観光訪問先の属性別分析, DEIM フォーラム 2015 最終論文集。