Flickr のジオタグ付き写真データからみる日本全土の観光特性

居住国推定とタグクラスタリングによる訪日外国人の興味分析

TOURISM CHARACTERISTICS THROUGHOUT JAPAN DETERMINED
FROM GEOTAGGED FLICKR PHOTO DATA

Interest analysis of foreign tourists based on estimated place of residence and tag clustering

北村 武士^{*1}, 本間 健太郎^{*2}, 今井 公太郎^{*3}

Takeshi KITAMURA, Kentaro HONMA and Kotaro IMAI

The purpose of this paper is to analyse the behaviour of foreign tourists using tags and geotags attached to Flickr photos. In this study, we analyse the spatio-temporal characteristics of photography related behaviour of Flickr users throughout Japan and estimate their places of residence. It is expected that such multivariate analysis of tourist behaviour will show differences in the 'attraction' (i.e., the unique charm that can be found only at that particular place) of tourist sites, as expected by the visitors from different places of residence.

Keywords : Flickr, Geotag, Tag, Estimation of places of residence, Multivariate analysis

フリッカー, ジオタグ, タグ, 居住地推定, 多変量解析

1. 序論

1.1. 研究の背景と目的

本稿の目的は、画像系 SNS として普及している Flickr のジオタグ付き写真データを用いて訪日外国人観光客の撮影行動を分析し、地域の観光特性を詳細に示すことである。

近年、観光時に撮影した写真は SNS を通して他のユーザと共有される機会が多くなっている。蓄積された写真は多くの情報価値を有しており、地域の観光特性を表しているものが多い。本稿では画像情報そのものは分析対象にしていなかったが、写真に付与されたジオタグやタグを用いて、世界各国の観光客の動態を解明する。

ジオタグとは、一般的に GPS 内蔵のデジタルカメラやスマートフォンでの撮影時に記録される、撮影場所の位置情報である。本稿では、ジオタグから撮影場所を特定するとともにユーザの居住国を推定する。

タグとは写真内容を簡潔に説明するキーワードであり、ユーザが写真にタグを付ける行為は近年一般的である。本稿では、ユーザが着目した撮影場所の魅力、すなわち地域の観光特性の一端を、タグがよく表していると捉えている。ユーザが独自の観点から写真に付与したタグは新たな魅力と成りえ、「実空間」の魅力を「情報空間」から日々更新させていると言える。

ユーザが発見した観光資源をタグとして大量に拾い上げる本研究は、観光立国を目指す日本において観光戦略立案に役立つだけでなく、地区スケールでのマーケティング、商業施設計画にも応用可能

であると考えている^{注1)}。

1.2. 先行研究

本稿は以下の新規性がある。(a)Flickr の投稿写真に付与されたジオタグを用いて、日本全土を対象とする広域での観光行動を分析している点。(b)約 17 年間にわたる写真データを取得し、観光行動の季節変動や時間帯変化を分析している点。(c)ユーザの居住国を推定し、居住国ごとに観光行動を分析している点。(d)写真に付与されたタグから地域の観光特性を分析している点。以上 4 点の新規性と先行研究を照らし合わせ、以下に本研究を位置付ける。

根舛(2014)¹⁾ は、観光行動は何によって規定されるかを明らかにするため、Flickr に投稿された写真を用いて、観光者の都市空間の「浸透度合い」を撮影場所の軌跡から可視化している。根舛らは、観光行動を都市スケールで分析しているが、本稿では日本全土を対象としている(新規性(a))。また、緒方(2013)²⁾ は、人が風景をたどる視線を Flickr に投稿された写真を用いて可視化することで、街路空間における魅力的な場所とそのつながりを分析している。緒方らは撮影対象を目視で判断しているが、本研究では写真に付与されたタグから機械的かつ統計的に判断している。このアプローチにより、撮影対象同定の精度は下がるものの、日本全国をシステムティックな分析を可能としている(新規性(a))。

倉田ら(2015)³⁾ は、時間帯別観光ポテンシャルマップ作成のために、Flickr 写真データ時刻補正と選別の方法を述べているが、実際に時間帯別の観光行動を分析するまでは至っていない。時空間分析

^{*1} 鹿島建設㈱^{*2} 東京大学空間情報科学研究センター 講師・博士(工学)^{*3} 東京大学生産技術研究所 教授・博士(工学)

Kajima Corporation

Lect., Center for Spatial Information Science, The Univ. of Tokyo, Dr.Eng.

Prof. Inst. of Industrial Science, The Univ. of Tokyo, Dr.Eng.

を行っている研究は少なく、その中に太田ら(2017)⁴⁾の研究がある。ここでは twitter のジオタグ付きツイートを対象にユーザーの特性を考慮した時空間分析を行っているが、データ取得期間が約 1 ヶ月間のため、季節によるユーザーの行動の違いまでは分析していない。本稿では、約 17 年間もの投稿写真データを使用するという特徴がある(新規性(b))。

Rugna ら(2012)⁵⁾は、ある都市に訪れた旅行者によって撮影された Flickr のジオタグ付き写真を集積し、旅行履歴からユーザーの居住国を特定する分析手法を生み出しており、その後の多くの研究で参考にされている。本稿においても Rugna らを参考にし、改良を加えた独自の方法でユーザーの居住国推定を行っている。また Rugna らは、推定した居住国のユーザーごとに観光行動を分析するまでは至っていないが、本稿ではそれを行っている(新規性(c))。

大森ら(2015)⁶⁾は、Flickr のジオタグ付き写真から抽出した海岸線に関する内容のタグを地図上にプロットし、実際の地形との一致度合いを数理的に評価している。Gentile ら(2011)⁷⁾は、Flickr のジオタグ付き写真から観光地に関するタグを抽出し、類似する観光地を割り出す手法を提案している。タグの内容に着目したこれらの研究は、山地や海岸線などの地理的特徴を直接的に表すタグのみに焦点を当てて分析している。それに対し本稿では、場所の持つ魅力に関するタグを網羅的に抽出し、地域の観光特性を分析している(新規性(d))。さらに新規性(c)と(d)を組み合わせ、観光時に求める地域の観光特性を居住国ごとに明らかにしている。

1.3. 本稿の構成

2 章では、Flickr API を使用^{注2)}し、Flickr のジオタグ付き写真データを得られる限り全て取得する。また、取得した写真データについて時空間特性を概観する。3 章では、撮影パターンからユーザーの居住国を推定する。また、調査データと比較することで居住国推定の精度確認を行う。4 章では、観光客の居住国の違いによる都道府県の選好の差異を分析する。また、ユーザーの居住国を都道府県の選好類似度から分類し、居住国クラスターを定義する。そして、居住国クラスターごとに撮影の地理的分布を分析する。5 章では、タグをその共起度にもとづき分類し、タグクラスターを定義する。また、タグクラスターごとに地理的分布を分析する。6 章では、居住国クラスターごとにタグクラスターの分布を分析し、観光時に求める地域の観光特性を明らかにする。また、都道府県スケールの観光特性を、観光客の居住国とタグクラスターの観点から示す。最後に 7 章で、本稿で得られた知見をまとめ、今後の課題と展望を述べる。

2. ジオタグ付き写真データの取得と概観

本研究では FlickrAPI を使用し、撮影期間が 2000 年 1 月 1 日 00:00:00 から 2016 年 10 月 31 日 23:59:59 の約 17 年間の写真を対象に、取得可能な Flickr の全ジオタグ付き写真データを取得した^{注3)}。本稿の分析対象は日本国内であるが、訪日客の居住国を推定するため、全世界での投稿写真を収集した。自動取得プログラムを 6 台のパソコンで 2 つずつ計 12 個実行し、2016 年 10 月 26 日から 2016 年 11 月 9 日の 15 日間の取得期間を要した。

取得写真データは計 287,471,240 件であった。約 3 億件の Flickr の写真データを取得している既往研究はなく、本研究の特徴のひとつである。

その後、取得写真から場所 ID を全て抽出し、FlickrAPI を使用することで場所 ID に対応する地名を取得し、撮影場所の国名や都道府県名を特定した^{注4)}。

その結果、アメリカで撮影された写真が最多であり、全体の約 30%に相当する計 87,738,082 件であった。日本で撮影された写真は 8 番目に多く、全体の約 3%に相当する計 8,695,914 件であった。

本稿では、日本で撮影されたこの約 870 万件の写真を分析対象とする。まず写真の地理的、時系列的、時空間的特性を概観する。

地理的特性: Fig.1 に日本全土における写真分布を示す。都市部での撮影が多く、山岳部では少ない。ただし山間部でも登山ルートでは比較的撮影が多く、特に富士山においては地形が読み取れるほど顕著である。また幹線道路や鉄道路線と一致する撮影が多い。しかし都市部から離れるに従い、その傾向が目立たなくなる一方で、駅・公園・観光名所などでは集中した撮影が目立つ。東京においては、公園・商業地域・観光地がある駅周辺での分布が目立つ。京都においては、基盤目状道路と一致する分布が確認でき、京都駅・二条城・京都御所・嵐山・清水寺・伏見稲荷大社といった観光地での分布が特に目立つ。また、東京と京都のいずれにおいても住宅地における撮影はあまり確認されなかった。

Fig.1 Distribution of photographed locations

時系列特性: Fig.2 に日本で撮影された写真枚数の 1 ヶ月ごとの変化を示す。4、11 月に多く、1、6 月に少ないという季節変動が見られた。また 2011 年は他年に比べて写真総数が少なく、これは東日本大震災の影響で観光客が減少したことが原因と考えられる。Fig.3 に平日と休日^{注5)}の 1 日における 30 分単位の平均撮影数を示す。平日と休日で夜朝の傾向に大きな違いは見られないが、6:00 以降から徐々に休日が平日より多くなり、日中を通して 2 倍程度になっている。撮影行動には余暇時間を利用した行動が表れており、観光行動を多く含むことが窺える。

Fig.2 Timeline of photo counts (Jan,2010-Dec,2015)

Fig.3 Average photo counts per 30-min interval

時空間特性: Fig.4に夏と冬における撮影場所の分布を示す^{注6)}。冬は夏よりも山間部での撮影が少なく、特に富士山頂上や槍ヶ岳、剣岳、薬師岳での撮影が極端に減少している。また、新潟県柏崎の海水浴場や栃木県の渡良瀬遊水地といった夏の観光資源が豊富な場所においても、同様に夏より冬の撮影が少なくない。

Fig.4 Seasonal distribution of photographed locations (Left: summer, Right: winter)

Fig.5に休日における日中と朝夜の撮影場所の分布を示す^{注7)}。日中は全体的に撮影が活発で、都心部に集中している一方で郊外での撮影も確認できる。朝夜は比較的撮影が少なく、都心部の繁華街に集中している。日中は代官山など商業圏、六義園など庭園、浅草など観光スポットにおける撮影が多く見られたが、夜朝は東京駅など主要駅周辺に集中しており、施設の営業時間との関係が窺える^{注8)}。

3. ユーザの居住国推定とその精度確認

本章では、まず日本で写真撮影したユーザを抽出し、次に該当す

Fig.5 Distribution of the photographs taken in Tokyo on holidays (Top: day-time, Bottom: night-time)

るユーザの各国での撮影日数をもとに統計的手法で居住国を推定している。また、居住国推定の精度確認を行うために観光庁の統計データと比較した。

3.1. 日本で写真撮影したユーザの抽出

2章で分析対象とした撮影場所が日本の写真を対象とし、ユーザIDをすべて抽出して重複を取り除いた結果、69,603ユーザが該当した。このうち、写真総数が30枚未満、全撮影日数が5日未満のいずれかを満たすユーザはそもそもFlickrをあまり活用していないユーザと判断し、以上いずれかの条件を満たすユーザを取り除いた37,212ユーザを分析の対象とする。

3.2. ユーザの居住国推定方法

居住国推定の方法は下記の順による。(1)ユーザごとに全ての投稿写真データを抽出する。(2)(1)で得た写真データについて、場所IDから撮影場所の国名、都市名を導く。(3)ユーザごとに各国の撮影日数^{注9)}を算出し、撮影日数が最も多い国を居住国とする。(4)居住国での撮影日数を全撮影日数で除した値を居住国滞在率とし、ユーザごとに居住国滞在率が50%以下のユーザは居住国推定の精度が低いと判断して取り除く^{注10)}。(5)(4)までで日本人と推定したユーザについては、居住都道府県も同様の手順で推定した。このとき居住都道府県滞在率^{注11)}が100%のユーザについては、投稿数が上位20ユーザを対象に写真の撮影対象を確認したところ、広告目的あるいは特定の地区で天体観測など特定の対象のみを撮影した写真がほとんどであった。これらのユーザは本稿の分析には適さないと判断し、

居住都道府県滞在率が 100%のユーザはすべて取り除いている。

以上の手順を踏み、居住国推定の精度が高いと判断された 26,245 ユーザを分析対象とした。

3.3. 都道府県訪問率データとの比較

対象のユーザの居住国を抽出したところ 145 カ国が得られた。このうち日本への訪問日数が 2000 日以上、14 カ国を対象とし、(a)観光庁が発行している統計データ^{注12)}と(b)居住国推定の結果と比較する。(a)(b)それぞれの 47 都道府県への訪問人数を導出し、そのうえで、居住国ごとに(a)と(b)の相関を分析した結果、12 カ国で相関係数が約 0.8 の強い相関が確認できた。Fig.6 に日本へ訪問人数が多い 3 カ国と最も相関係数が低い韓国の相関関係を示す。韓国は相関係数が約 0.6 と他の居住国より低いが、これは「福岡」「大分」など頻繁に訪れている都道府県では撮影が減少することが原因と考えられる。Flickr データと統計データの強い相関から Flickr データは実際の外国人観光客の動態をよく表していると言え、また居住国推定の精度が高いことが確認された。

Fig.6 Confirmation of the accuracy of the estimated place of residence

4. 居住国クラスターと撮影場所の関係性

本章では、まず分析の対象とする観光客を分類する。次に観光客の居住国による都道府県選好の違いを分析する。また、居住国を都道府県の選好類似度からクラスタリングし、居住国クラスターを定義する。最後に居住国クラスター別に撮影場所の分布を示し、居住国による観光行動の差異を分析する。

4.1. 観光客の分類

前章で推定したユーザを非観光客、日本人観光客、訪日外国人観光客の計 3 つに分類する。日本人ユーザに関しては、撮影場所が居住都道府県内に含まれる場合は非観光客とし、含まれない場合は日本人観光客として扱っている。訪日外国人観光客は、前章で居住国推定の精度確認を行った 14 カ国を対象とし、アメリカについては

西経 100 度で西部と東部で分割して分析する。この計 15 の居住国（便宜的に「国」と呼ぶ）の訪日外国人観光客および日本人観光客を分析対象とする。

4.2. 選好度の定義

本節では、選好度 X_{ij} (居住国 i の観光客が都道府県 j を選好する相対的な割合) を定義する。純粋な居住国の違いによる都道府県の選好の違いを分析するためには、居住国ごとに異なる人口や Flickr 利用率、都道府県の規模による訪問人数の多寡を標準化する必要がある。そのために、まず観光客の居住国別に各都道府県での撮影日数 D_{ij} を求める。次に居住国 i のユーザによる総撮影日数 ($\sum_j D_{ij}$) に、都道府県 j で撮影される「確率」($\sum_i D_{ij} / \sum_i \sum_j D_{ij}$) を乗じ、式(2)に示す E_{ij} を求める。この E_{ij} は、居住国 i の観光客が都道府県 j で撮影する日数の「理論値」と言える。その「理論値」と観測値 D_{ij} の相対誤差を選好度 X_{ij} と呼ぶ (式(1))。

$$X_{ij} = \frac{D_{ij} - E_{ij}}{E_{ij}} \quad (1)$$

ただし、

$$E_{ij} = \frac{\sum_i D_{ij} \sum_j D_{ij}}{\sum_i \sum_j D_{ij}} \quad (2)$$

D_{ij} : 居住国 i の観光客が都道府県 j で写真を撮影した日数

選好度 X_{ij} の数値をヒートマップで示す(Fig.7 中央)。Fig.7 を見ると、例えば日本人観光客(i =Japanese Tourist)による(j =Tokyo)への選好度は際立って低い。この事実は、「日本人観光客が東京に観光に行かない」ことを示しているのではなく、また「他の都道府県に比べて東京に行かない」ことを示しているのでもない。そうではなく、『他の都道府県に比べて東京に行く度合い』が、他の居住国の観光客に比べて日本人は低い」ことを示している。この 2 重の相対性によって、式(1)で定義される選好度 X_{ij} は、「居住国 i からの来日客数」と「都道府県 j の観光客数」の両方の影響をキャンセルしている。

4.3. 居住国および都道府県のクラスタリング

本項では、前項で定義した選好度 X_{ij} を用いて居住国と都道府県のクラスタリングを行う。居住国 k と l の非類似度を式(3)に示す $u[k, l]$ とし、都道府県 m と n の非類似度を式(4)に示す $v[m, u]$ と定義する。

$$u[k, l] = \sqrt{\sum_j (X_{kj} - X_{lj})^2} \quad (3)$$

$$v[m, u] = \sqrt{\sum_i (X_{im} - X_{in})^2} \quad (4)$$

これを距離指標として用い、word 法としたクラスタリングの方法でクラスタ分析を行い、Fig.7 の上部に居住国、左部に都道府県のデンドログラムを描いた。Fig.7 は、得られた居住国と都道府県のデンドログラムに従い選好度 X_{ij} を並び替えている。

4.4. 居住国による都道府県選好度の差異

Fig.7 のヒートマップから、日本人観光客は訪日外国人観光客があまり選好しない都道府県においても写真撮影を行っており、訪日外国人とは全く異なる都道府県を選好していることが明らかである。また訪日外国人の中でも居住国が欧州の観光客は、他国よりも東京・京都・広島を選好し、アジアの観光客は大阪・北海道・沖縄を選好することが分かる。また、ドイツの観光客は神奈川・千葉・茨

城・兵庫といった東京および京都の近隣都道府県を愛好しており、他の欧米諸国と比較して独特な観光行動をとっていることが分かる。

また、アメリカ西部からの観光客は青森、東部からの観光客は栃木をより愛好しており、局所的な選好の差異も明らかになった。

Fig.7 Clustering and heat maps of places of residence and prefecture

4.5. 選好度による居住国と都道府県のクラスタリング

Fig.7 上部における居住国のデンドログラムから、観光客の居住国について3つのクラスターを抽出した。各居住国クラスターを「日本人」「欧米人」「アジア人」と呼ぶ。「欧米人」は欧米諸国、「アジア人」はアジア諸国をすべて含んでおり、都道府県の選好パターンのみからこのように明確な居住国クラスターに分類されたことは、地理的文化的に接近した観光客は似た地域を訪れていることを示している。

Fig.7 左部における都道府県のデンドログラムから、都道府県をA、B、C、Dの4つのクラスターに分類した。クラスターA(東京、京都、広島)は欧米人が選好しやすく、クラスターB(大阪、北海道、沖縄、福岡)はアジア人が選好しやすいことがわかる。クラスターC(栃木、奈良、愛知、石川など)はイタリア人、スペイン人、フランス人が特に選好しやすく、クラスターD(茨城、大分、新潟、島根など)は日本人観光客が選好しやすい都道府県であると分かった。

4.6. 居住国クラスター別にみた撮影場所の分布

前節より、京都は欧米人が、大阪はアジア人がより選好しやすいことが明らかになった。この上で、京都と大阪についてアジア人と欧米人の撮影場所の分布を示す(Fig.8)。アジア人は欧米人よりも、京都の嵯峨野観光鉄道に沿った撮影が特に目立ち、大阪では桜之宮公園、帆船型観光船サンタマリアの航路などにおいて、線状に撮影場所が分布している。また、アジア人は甲子園や花博記念公園鶴見緑地など、欧米人の分布があまり見られない場所においても撮影が集中していることが分かった。これらについては、アジア人と欧米人の投稿数の多寡を考慮したとしても、分布に顕著な違いがあることが観測できる。全体的には、撮影場所がアジア人は凝集的であることに對し、欧米人は分散的であることが読み取れる。このことから、アジア人は観光先が一致しがちなのに対し、欧米人は観光行動に個人差があるまたは個々が分散的な観光をするといった、居住国によって観光行動の特徴が異なることが推察される。

5. タグ属性と撮影場所の関係性

本章では、ユーザが写真に付与したタグについて分析する。まず、写真に付与されたタグの内容と投稿数について概観する。次に、地域の観光特性を表すタグのみを抽出し、タグの類似度を統計的手法で算出し、クラスタリングする。また、クラスターの内容をもとにタグクラスターを定義する。最後にタグクラスターの分布を確認し、地理的特性との一致を確認する。

5.1. タグの抽出

日本で撮影された全ての写真の約71.1%に相当する6,183,103件の写真にタグが付与されていた。それらのタグの種類は、917,716種であった。Table1にタグの内容を、それが付与された写真の数とともに、多い順に示す。タグの上位は、“japan”“tokyo”などの地名や“square”“nikon”などカメラ情報のタグが占めていた。Fig.9は、各タグが付与された写真数を縦軸に、それに基づいた各タグの順位を横軸に取った散布図である(両対数グラフ)。右肩下がりの直線に従っているといえ、ユーザが写真に付与させるタグはランキング上位に集中する一方で、1枚の写真にしか付与されないタグも多い。

5.2. タグの分類

タグをタグ間の類似性から分類する。本稿では地域の観光特性を分析するため、場所を直接示す地名や季節・日時などの時間的情報、カメラ機種、人名やグループ名、特定のユーザが大量に付けたタグなどを取り除いたうえで、投稿数が上位 100 個のタグを分析対象とする^{注 13)}。しかし、似た意味を持つタグが複数存在するので、分かりやすい分析結果を得るためそれらをまとめたい。そこでまず、タグ i とタグ j の非類似度を表す値として式(5)に示す Jaccard 距離 w_{ij}

Fig.8 Distribution of the locations photographed by Asians and Westerners (1, 2: Kyoto; 3, 4: Osaka)

Table1 Number of photos that have the top-22 tags

Tag	Number of photos	Tag	Number of photos
japan	2,478,100	geotagged	178,750
tokyo	944,417	東京	161,117
日本	524,538	osaka	140,171
square	429,124	canon	138,944
iphonephotography	428,525	jp	134,641
squareformat	423,801	japanese	125,860
instagramapp	422,788	asia	124,151
upload:by=instagram	421,674	eyefi	122,932
kyoto	354,511	japon	117,569
nikon	234,862	hokkaido	110,927
travel	218,993	temple	104,700

Fig.9 Double-natural-logarithmic graph

【Shrine&Traditional】 【神社・伝統】 shrine shinto torii traditional culture matsuri kimono	【Temple&Buddhism】 【寺院&仏教】 temple buddha statue buddhism buddhist	【City】 【都市】 town city urban road street view cityscape architecture building
【Nature (green)】 【自然 (緑)】 nature sky blue bike mountain tree green landscape	【Waterside】 【水辺】 sea beach ocean island river water bridge	
【Sakura&Hanami】 【桜&花見】 cherry sakura blossom plant hanami cherryblossom backpackers	【Disaster】 【災害】 earthquake tsunami	
【Live concert】 【音楽ライブ】 live music rock	【Tourism&Event】 【観光&イベント】 aquarium museum art garden plants rose park animal bird zoo illumination scenery tower buildings restaurant heritage castle house sign high outdoor bicycle lake snow baseball minato market hotel onsen event wedding pokemon show car color red white pink cute	
【Cosplay】 【コスプレ】 portrait woman cosplay cosplayer	【Aviation】 【航空】 airport airplane	【Railway】 【鉄道】 train rail station railway
		【Black-and-White Photograph】 【モノクロ写真】 blackandwhite monochrome

Fig.10 Classification of tags via the k-means method (13 clusters)

を算出する。この w_{ij} は、タグの共起しにくさ(同じ写真のタグになりにくさ)を意味している。

$$w_{ij} = 1 - \frac{\text{タグ}i\text{とタグ}j\text{の両方を持っている写真数}}{\text{タグ}i\text{またはタグ}j\text{のどちらかを持っている写真数}} \quad (5)$$

次に、 w_{ij} を用いて k-means 法によるクラスタリングからタグを分類する。k-means 法のクラスター数を 2 個から順に増やすにつれ、「コスプレ」「桜・花見」「災害」の順に特定のイベントが分離された。5 個からは、おおよそ「都市」「自然」「宗教」などが分離された。最終的には各クラスターの内容の意味づけが可能な 13 クラスターで分割を終了した。各クラスターを、その内容から適切な名称を付けたうえで、タグクラスターと呼ぶことにした(Fig.10)。「桜・花見」タグクラスターの中に「backpackers」のタグが見られるなど意外な組み合わせも見つかった。色に関するタグについては、ほとんどは「観光&イベント」に分類されたが、「blue」「green」については「自然(緑)」に強く関係していることがわかる。

5.3. タグクラスター別にみた撮影場所の分布

前節で分類したタグクラスターと撮影場所の分布を概観する。Fig.11 にタグクラスター「水辺」「自然(緑)」「鉄道」「災害」が付与された写真の分布を示す。互いのタグクラスターと比較したとき、全体的な投稿数の多寡を考慮しても分布に顕著な違いが確認でき、以下の特徴があった。「水辺」は、海岸線上や離島に多く分布している。「自然(緑)」は、富士山や甲武信ヶ岳など山間部に集中している。「水辺」「自然(緑)」は意外にも東京都心部に集中していたが、川や池、都市公園や街路樹などを撮影した写真に付与されたタグが原因であると考えられる。また、「鉄道」は都心からライン上に伸びて分布しており、実際の駅近辺に集中している。「災害」は東北地方沿岸部に集中しており、東日本大震災以降に付与されたタグが多い。こ

Fig.11 Distribution of tag clusters
(Upper left: Waterside, Upper right: Nature(green),
Lower left: Railway, Lower right: Disaster)

のように分類したタグクラスターと実際の地理的特徴との一致が確認できた。

6. 居住国クラスター別にみた地域の観光特性の分析

本章では、4 章で分類した「アジア人」と「欧米人」について、5 章で分類したタグクラスターの分布を比較する。都市スケールと地区スケールのそれぞれ同地区でタグクラスターの分布を比較し、地域の観光特性を分析する。また、47 都道府県ごとに持つ居住国クラスターとタグクラスターを組み合わせた成分を、主成分分析によって 2 成分に縮約することで、都道府県の観光特性を定量的に比較できる指標を作成する。

6.1. 都市スケールの観光特性の分析

本節では、都市スケールとして東京都心部を対象に、「アジア人」と「欧米人」が写真に付与した「観光&イベント」タグクラスターの分布を比較する(Fig.12)。同じタグクラスターでも、ユーザの居

Fig.12 Distribution of the tag cluster 'Tourism & Events' in Tokyo
(Top: Asians, Bottom: Westerners)

住国ごとに撮影場所の違いが分布に明確に表れており、これは居住国による投稿数の違いを考慮しても顕著な分布の差である。相対的に「アジア人」はお台場やホテル椿山荘で多く、「欧米人」は歌舞伎町や渋谷駅近辺、皇居、銀座近辺、観光汽船で多いことから、実際の観光行動をよく表していると言える。

6.2. 地区スケールの観光特性の分析

本節では、6.1 よりも詳細な観光行動を明らかにするために、新宿を対象とした地区スケールの分析を行う。Fig.13 にタグクラスター「桜&花見」「観光&イベント」についての「アジア人」と「欧米人」の分布を示す。「桜&花見」は「アジア人」のほうが多くつけていることから、特に興味を示していることが読み取れ、特に新宿御苑や新宿二丁目、新宿中央公園での分布が目立つ。「観光&イベント」は「欧米人」のほうが多くつけており、特に歌舞伎町での分布が目立つ。また「欧米人」のほうが広範囲かつ分散的に分布しており、このことから4章で示した「欧米人」のほうが観光行動に個人差が見られることについて、タグの観点からも再度確認できた。

Fig.13 Distribution of tags by Asians and Westerners in Shinjuku (Top: Sakura & Hanami, Bottom: Tourism & Events)

6.3. 都道府県スケールの観光特性の分析

本節では、どの居住国の観光客が、どの地域の観光特性をタグとして写真に残したかという観点のもとに、都道府県のポテンシャルを該当する写真数から定量化し、これを都道府県の観光特性として分析する。観光特性については、5章で分類したタグクラスターを「自然」「都市」「歴史」を3つに再分類したものを対象とする^{注14)}。

具体的には、上記の3分類されたタグごとに、4.2.で示した方法によって選好度を算出した。ただし居住国*i*の変数として、16居住国ではなく、4章で分類した「日本人」「アジア人」「欧米人」の3つの居住国クラスターを用いた。また撮影日数 D_{ij} でなく、下式に記すように撮影枚数 D_{ijk} を用いている。以上の手続きを数式化すると選好度 X_{ijk} は下式で表される。

$$X_{ijk} = \frac{D_{ijk} - E_{ijk}}{E_{ijk}} \quad (6)$$

ただし、

$$E_{ijk} = \frac{\sum_i D_{ijk} \sum_j D_{ijk}}{\sum_i \sum_j D_{ijk}} \quad (7)$$

D_{ijk} : 居住国クラスター*i*の観光客が都道府県*j*で撮影したタグ*k*の写真数

都道府県*j*ごとに、選好度 X_{ijk} は居住国クラスター($i=1,2,3$)×タグ($k=1,2,3$)の計9つ存在している。この厳選された9次元の属性ベクトルは、都道府県の観光特性をよく表しているといえる。さらに次元を縮約するために47都道府県×9属性のマトリックスに対して主成分分析を行った。Fig.14に第1主成分を横軸と第2主成分を縦軸とした9属性の散布図を示す。第1主成分は「日本人・自然」「アジア人・自然」が高く、「欧米人・歴史」「日本人・歴史」「アジア人・歴史」が低いことから「自然観光資源度(natural attractions)」と解釈できる。第2主成分は「日本人・自然」「日本人・歴史」が高く、「アジア人・自然」「アジア人・都市」が低いことから「国内観光客度(attractions of Japanese tourist)」とした。同じ縦軸・横軸に対し、47都道府県をプロットした散布図をFig.15に示す。これは各都道府県の観光特性を総体的に評価した図である。沖縄、広島は、相対的に自然観光資源を外国人観光客に好まれており、京都、東京、大阪、福岡は、歴史観光資源が外国人観光客により好まれている。新潟、富山、福島、鹿児島、徳島、山梨は、いずれも自然観光資源は同程度だが、新潟や富山のほうが徳島や山梨よりも日本人に好まれている。都道府県によって異なる魅力を、「自然観光資源度」と「国内観光客度」の観点から都道府県スケールの観光特性として位置付け、都道府県ごとの魅力を定量的に比較することが可能となった。

Fig.14 Scatterplot of the attribute vectors clustering around the first and second principal components

Fig.15 Scatterplot to assess the overall 'attraction' of individual prefectures

7. 総括と展望

本研究は、Flickr のジオタグ付き写真から観光行動を分析し、観光客の居住国とタグの観点から地域の観光特性を記述した。

2 章では、Flickr の全ジオタグ付き写真データを取得し、日本全体での撮影行動の時空間特性を分析し、夏冬や昼夜における撮影場所の違いなどを明らかにした。

3 章では、ユーザごとの世界各国における撮影日数の多寡から居住国推定を行った。また、Flickr データと都道府県訪問率データの強い相関も確認され、居住国推定の精度が高いことも裏付けされた。

4 章では、日本人観光客及び 15 居住国の観光客について、都道府県の選好度を算出した。この選好度をもとに居住国をクラスタリングした結果、日本人、アジア人、欧米人の 3 つに明確に分類された。これは、地理的文化的に接近した観光客は似た場所に行くことを示している。またヒートマップや撮影場所の分布から、アジア人は観光地近辺に撮影が集中するのに対し、欧米人は観光地から少し離れた場所では分散的な撮影行動をする差があることが確認でき、居住国による観光行動の違いが明らかになった。

5 章では、写真のタグと撮影場所の関係を分析した。タグを、その非共起度から k-means 法でクラスタリングしたところ、「水辺」「鉄道」などの 13 タグクラスターに分類され、その分布は実際の場所との整合性も確認できた。

6 章では、アジア人と欧米人について、特定のタグクラスターの分布を都市スケールと地区スケールで比較した。これは実際の観光行動をよく表しており、居住国ごとの地域の観光特性に違いがあることを確認できた。また、アジア人については、欧米人に比して限られた観光地に人気が集まるということも分かった。その上で、居住国とタグを組み合わせた 9 次元の属性ベクトルを主成分分析し、都道府県スケールでの観光特性を「自然観光資源度」と「国内観光客度」で定量的に比較できる指標を示した。本稿では、都道府県スケールで観光特性を評価しているが、これは地区・街区レベルで評価するには現状の写真データ数では不十分であると判断したためである。将来的に写真データ数が増えれば、より詳細なレベルで「自然観光資源度」と「国内観光客度」による評価が可能となり、観光資源およびターゲットを考慮した観光計画にも展開できる。

本稿の研究手法は、他の SNS などのビックデータにおいても有

効であると言える。また、本研究では扱わなかったユーザの投稿履歴をたどり、タグを用いて観光地の魅力変数を設定することで、居住国別の観光地選択行動モデルを作成できると展望する。

謝辞

本研究は JSPS 科研費 16K20971 (「大規模な位置情報データを用いた観光行動の精緻モデリングによるツーリズムデザイン」, 研究代表者: 中間健太郎) の助成を受けたものです。

参考文献

- 1) Nemasu, T.: Kankoukoudounimirutoshinopamiabirityi (Permeability of cities in terms of tourism behavior), Master's thesis collection of Dept. of Arch., The Univ. of Tokyo in 2014, 2014 (in Japanese)
根岸卓也: 観光行動にみる都市のパーミヤビリティ, 東京大学大学院建築学専攻2014年度修士論文梗概集, 2014
- 2) Ogata Y.: Toshikukanniokerushitenbatoshitaisyonokenkyu Flickr API wo mochiitashisenbunnokashikatobunseki (A study of visual point and visual object in urban space -Visualization and analysis of line of sight using Flickr API-), Master's thesis collection of Dept. of Arch., The Univ. of Tokyo in 2013, 2013 (in Japanese)
緒方祐磨: 都市空間における視点場と視対象の研究 Flickr Apiを用いた視線分の可視化と分析, 東京大学大学院建築学専攻2013年度修士論文梗概集, 2013
- 3) Kurata Y. and So N., Ishikawa H., Sanada H., Ikeda T.: Jikantaibetsu kankopotensyarusakuseinotamenoFlickrsyashindetanajikokuhoseitosenbetsu, DEIM Forum (Time correction and selection of Flickr photo data For tourism potential map by time zone), F7-6, 2015 (in Japanese)
倉田陽平, 相尚寿, 石川博, 真田風, 池田拓生: 時間帯別観光ポテンシャルマップ作成のためのFlickr写真データの時刻補正と選別, DEIM Forum, F7-6, 2015
- 4) Ota, K. and Imai, K., Honma, K.: A STUDY ON IDENTIFICATION OF REGIONAL CHARACTERISTICS BASED ON TEMPORAL-SPATIAL ANALYSIS OF GEOTAGGED-TWEET DATA, Journal of Architecture and Planning (Transactions of AIJ), Vol.82, No.731, pp.283-289, 2017.1 (in Japanese)
太田圭亮, 今井公太郎, 本間健太郎: ジオタグつきtweetの時空間解析に基づいた地域特性抽出に関する研究, 日本建築学会計画系論文集, 第82巻, 第731号, pp.283-289, 2017.1
- 5) Jérôme Da Rugna, Gaël Chareyron, Bérengère Branchet: Tourist behavior analysis through geotagged photographs: A method to identify the country of origin, 13th International Symposium on computational Intelligence and Informatics (CINTI), Budapest, Hungary, pp.20-22, 2012
- 6) Omori M. and Hirota M., Ishikawa H., Yokoyama S.: Extraction and Evaluation of Geographical Characterization Based on Geo-tags on Social Media, Transactions of Information Processing Society of Japan, Vol.8, No.1, 1-16, 2015.3
大森雅己, 廣田雅春, 石川博, 横山昌平: ソーシャルメディア上から収集したジオタグに基づく地理的特徴の抽出と評価, 情報処理学会論文誌, Vol.8, No.1, 1-16, 2015.3
- 7) Gentile, L.: Using Flickr Geotags to Find Similar Tourism Destination, Department of Computer Engineering at Politecnico di Milano, Milano, Italy, 2011

注

- 注 1) 筆者が写真の撮影対象とジオタグのずれを街区スケールで確認したところ、最大 50m 程度のずれが確認された。ジオタグの精度は撮影機器の GPS の精度に依存し、Flickr に投稿される写真の多くはスマートフォンで撮影されているため、スマートフォンの GPS の精度に大きく影響される。太田らの研究⁴⁾では、スマートフォンを利用し、tweet を投稿した位置とジオタグデータとの距離の差を計測した結果、精度誤差の平均は約 12m であったとの記述がある。本稿ではこのジオタグの精度を参考にし、街区スケールの分析をする上で、十分なジオタグの精度があると判断している。
- 注 2) FlickrAPI の利用規定では、研究目的でのデータ使用を禁止していない。なお商用目的の場合は審査がある。
- 注 3) 本稿では、FlickrAPI の flickr.photos.search を利用し、検索条件とし

てジオタグの有無や写真の撮影期間を設定し、写真データを取得している。同一の検索条件で取得できる写真データ数は最大 4000 件のため、写真の撮影期間を細かく分割することで検索条件にあてはまる写真数が 4000 件以下になるようにしている。すなわち撮影期間が 1 時間の範囲におさまるように検索条件を設定し、これに当てはまる写真が 4000 件以上ある場合は 4000 件以内に収まるように検索期間を更に分割して取得している。また、取得した写真データには重複が含まれることは避けられないため、データ取得後に重複を取り除いている。

注 4) 場所 ID とは、ジオタグの緯度経度情報をもとに地名を ID 化した記号であり、適切な API を使用することで、場所 ID から撮影場所の地名を特定できる。Flickr に投稿されたジオタグ付き写真には、Flickr 社と Yahoo(米)のそれぞれが別に定めた 2 種類の場所 ID が付与されている。本研究では、FlickrAPI の flickr.places.getInfo を利用することで、2 種類の場所 ID から "Kyoto-shi, Kyoto Prefecture, Japan" のような地名を得ている。既往研究では、Google Geocoding API などを使用した緯度経度情報の逆ジオコーディングが一般的であるが、この API 使用回数制限は 1 日当たり 2500 回であるため、約 3 億枚の写真データを取り扱う本研究では上記の方法を採用した。

注 5) 土日及び祝日を休日とし、それ以外を平日と定義した。

注 6) 気象庁による季節の区分を参考にし、春(3-5 月)、夏(6-8 月)、秋(9-11 月)、冬(12-2 月)としている。

注 7) 9 時から 21 時までを昼、21 時から 9 時までを夜朝としている。

注 8) 日本全国における 30 分単位の平均撮影数は、休日は約 3000 枚、休日朝夜は約 650 枚で約 5 倍の差がある。Fig.5 の例えばディズニールランドを見ると、昼の撮影数は朝夜に比べて 15 倍程度多い。これに対し、渋谷駅周辺 300m の範囲では、朝夜に対する昼の撮影数は 2 倍程度に過ぎない。このように、撮影時間と施設の営業時間には関係が窺える。

注 9) 写真撮影をした日は撮影場所の国と都市に 1 日滞在していたとし、全ての国と都市について滞在日の合計を滞在日数とした。

注 10) Rugna らの研究⁵⁾では、公開写真が 30 枚以上のユーザを対象とし、訪問国と訪問都市の滞在日数を順位付けし、「最も滞在日数が多い都市」および「最も滞在日数が多い国」を明らかにすることで、ユーザの居住国を判断している。本稿も既往研究に倣い、写真総数が 30 枚未満のユーザを除外した。加えて、全撮影日数が 5 日未満のユーザは、撮影場所が 1、2 箇所程度で居住国推定が困難なため、これも除外した。さらに、居住国滞在率が 50%以下のユーザについても、滞在日数が 2 番目に多い国との滞在率の差があまりなく、また最も滞在日数が多い国と最も滞在日数が多い都市が属する国が異なる場合が多く確認されたため、居住国推定の精度が低いと判断し、本稿の分析から除外した。

注 11) 居住都道府県滞在日数を居住国滞在日数で除した値である。

注 12) 観光庁の訪日外国人消費動向調査の都道府県訪問率[観光レジャー目的]: 2010 年度～2015 年度を使用。

注 13) タグの分析対象を選定する上で、京都のジオタグが付与された写真に "TOKYO" のタグが付与されているなど、ジオタグとタグの不一致が確認された。これらの写真の被写体から判断すると、ジオタグの位置情報が正しいケースが多いため、タグから地名を除いた場合、被写体とその他のタグの内容の不一致は少なくなる。このような被写体とタグの不一致は、全体の数%程度であり、また系統的な誤差ではなく個別の誤差であるため、本稿で用いる統計的な分析手法には影響を与えないと判断している。

注 14) 5 章で分類したタグクラスターのうち、[水辺][自然(緑)]を「自然」、[都市]を「都市」、[神社・伝統][寺院・仏教]を「歴史」と分類している。[観光・イベント]のクラスターについては、タグの内容が多様で複数の異なる観光特性が混ざっている状態であると判断したため取り除いている。「鉄道」を含むその他のタグクラスターについては、空間的偏在が少ないので、観光特性の分析には適さないと判断し、取り除いている。

TOURISM CHARACTERISTICS THROUGHOUT JAPAN DETERMINED FROM GEOTAGGED FLICKR PHOTO DATA

Interest analysis of foreign tourists based on estimated place of residence and tag clustering

Takeshi KITAMURA^{*1}, *Kentaro HONMA*^{*2} and *Kotaro IMAI*^{*3}

^{*1} Kajima Corporation

^{*2} Lect., Center for Spatial Information Science, The Univ. of Tokyo, Dr.Eng.

^{*3} Prof., Inst. of Industrial Science, The Univ. of Tokyo, Dr.Eng.

In this study, tourist behaviour was analysed based on geotagged Flickr photo data, and a description of the ‘attraction’ of a place was provided from the viewpoint of a tourist’s place of residence and their tags. This study is novel in the following ways: (a) it analyses tourist behaviour throughout Japan using the geotags attached to posted Flickr photos; (b) using the photographic data collected over about 17 years, it analyses changes in tourist behaviour depending on the season and the time of the day; (c) it analyses tourist behaviour based on users’ estimated place of residence and (d) the ‘attraction’ of the photographed locations is analysed from the tag attached to the photo.

This study is structured as follows, with the results presented below.

In Chapter 2, we acquired Flickr geotagged photo data and clarified the spatio-temporal characteristics of photography related behaviour throughout Japan.

In Chapter 3, we estimated the place of residence of each user based on the number of days they spent taking pictures in each country in the world. In addition, a strong correlation was confirmed between Flickr data and the rate of visits to each prefecture, and a high estimation accuracy of users’ places of residence was confirmed.

In Chapter 4, the prefecture preference was calculated for Japanese tourists as well as for tourists from 15 different countries. By preference-based clustering, the places of residence were clearly classified into three groups, namely Japanese, Asians and Westerners. Our analysis indicates with high accuracy that tourists from geographically and culturally similar countries visit similar places. Furthermore, from the distribution of heat maps and photographed places, we find differences in tourist behaviour depending on the places of residence such as Westerners visiting Kyoto.

In Chapter 5, we analysed the relationship between photo tags and photographed locations. The tags were clustered by the k-means method based on their degree of asynchrony and were classified into 13 tag clusters such as ‘waterside’ and ‘railroad’, and their data were confirmed to be consistent with that of the actual place.

In Chapter 6, we compared the distribution of specific tag clusters of Asians and Westerners at both city and town levels. This correctly represented the actual tourist behaviour, and we confirmed that the ‘attraction’ depended on the place of residence. We also found that Asians’ preferences concentrated on relatively limited tourist spots as compared with Westerners’ preferences. Principal component analysis was performed on nine-dimensional attribute vectors created by combining the place of residence and the tag, and we succeeded in finding quantitative indicators to help with comparing the ‘attraction’ of the prefectures using ‘natural attractions’ and ‘attractions of Japanese tourist’ as parameters.

The study method in this paper is also considered an effective application of big-data techniques to another SNS. In addition, we will be able to construct a tourist destination choice model applicable to every place of residence by tracing the posting history of users who were not handled in this study and setting the attractiveness variables of tourist spots using tags.

(2018 年 4 月 9 日原稿受理, 2018 年 9 月 25 日採用決定)