

ジオタグ付きソーシャルメディアを用いた国内観光資源の把握と インバウンド観光のギャップ分析

Extraction of Domestic Tourist Attraction and Analysis of Inbound Tourist's Visiting Places
Using Geotag Data from Social Media

前田 高志ニコラス *¹
Takashi Nicholas MAEDA

吉田 光男 *²
Mitsuo YOSHIDA

鳥海 不二夫 *¹
Fujio TORIUMI

大橋 弘忠 *¹
Hirotsuda OHASHI

*¹東京大学 大学院工学系研究科

Graduate School of Engineering, The University of Tokyo

*²豊橋技術科学大学 情報・知能工学系

Department of Computer Science and Engineering, Toyohashi University of Technology

The number of inbound tourists has been increasing recently in Japan. However most inbound tourists visit only certain areas. To balance regional economic growth in a country, it is necessary to reduce the spatial convergence of inbound tourists' visiting places. In order to reduce the spatial convergence of inbound tourists' visiting places, we have to know where tourist attractions are located and how different in preferences between foreign tourists and domestic tourists. In this research, we aim to investigate a method to show the difference of preferences between foreign tourists and domestic tourists. This research focuses on historic sites in Japan, but this framework can be applied to any other countries and any other kind of tourist attraction.

1. はじめに

日本を訪れる外国人観光客数は年々増加しており、政府および民間において今後も観光産業の活性化が望まれている。しかし、現状、日本を訪れる外国人観光客の訪問地が東京・京都・大阪といった一部の地域に集中していることが指摘されている。このため、日本のその他の地域に現存する観光資源に外国人観光客を引きつけることにより、観光による経済の活性化を、日本の広い地域にもたらすことが必要である。

本研究は、国内の観光資源の把握、および外国人観光客と国内観光客の訪問地の特性の違いを明らかにすることを目指す。これらにより、外国人観光客が観光地に求める施設の種類の違い、未だ外国人観光客が訪れていない観光地にそのような施設を備えることで、外国人観光客をそれらの観光地に引きつけることができると考える。

Vuら [Vu 15] は Flickr のジオタグ付き写真データから、香港における欧米からの旅行者とアジアからの旅行者について、人気の場所とルートと比較する研究を行っている。Paldinoら [Paldino 15] は同じく Flickr のジオタグ付き写真データから、アメリカ・ヨーロッパの各都市において、外国人観光客の訪問地の分布が国内観光客の分布よりも小さい範囲に集中していることを示している。しかし、これらの研究においてはなぜ観光客の出身地によって訪問地が異なるかを説明していない。これを説明するためには、各地域内の施設に関する情報を利用する必要がある。Georgievら [Georgiev 14] は、ロンドンオリンピックにおいて、どのような施設が訪問客数を顕著に伸ばしているかを分析している。ここで、Foursquare のデータが用いられている。Foursquare は各施設のカテゴリ（ホテル、レストランなど）を保持しており、その情報が利用されている。

本研究は、2014年8月に取得したジオタグ付き Tweet と Foursquare のデータから日本の史跡を持つ観光地について、

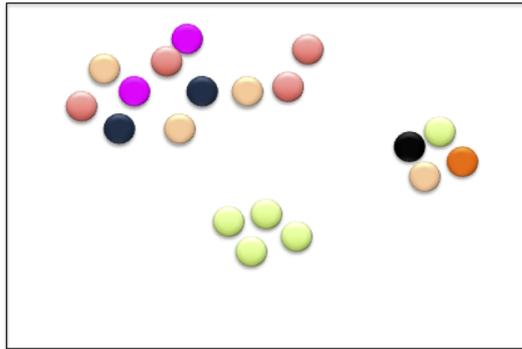
国内旅行者と外国人観光客の訪問しやすい地域の特徴の違いを分析する。両者の訪問地の特徴の違いを明らかにすることで、外国人観光客が訪れていない国内の観光資源に彼らを引きつけるために必要な対策案を提示する。

2. 国内居住者の訪問地抽出

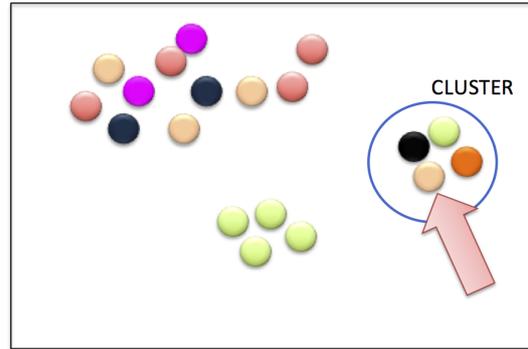
国内居住者の居住地と訪問地の推定・分類を Esterら [Ester 96] によって提案された DBSCAN (Density Based Spatial Clustering Algorithm with Noise) を用いて行う。DBSCAN とは、人の移動ログが集中する場所はその人にとっての日常の重要な場所（職場・家など）であるという仮定のもと、ログの空間上の密度をもとに、重要な場所を CLUSTER とし、そうでない場所を NOISE とみなすアルゴリズムである。

ジオタグ付き Tweet は、ユーザが能動的に投稿するものであるため、単に投稿地点の密集地を捉えてしまうと、一時的に多数の投稿をしただけの場所を日常の重要な場所と捉えてしまいかねない。これを避けるために、投稿日数の密度を考慮したアルゴリズムに改良する。以下に本研究で用いたアルゴリズムの詳細を示す。

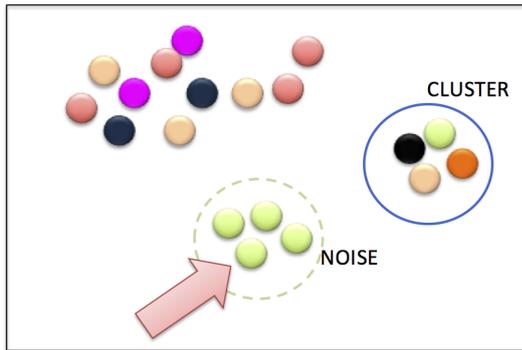
1. 1人のユーザに関し、期間内の全 Tweet の位置を抽出する。図 1-(a) は得られた Tweet の空間的な位置を示す。色の違いは投稿日の違いを示す。
2. この中からどれか 1 点を取り出し、半径 4km 以内に 4 日分以上の Tweet があれば、それらを同じクラスタとみなす (図 1-(b))。
3. 半径 4km 以内に 4 日分以上の Tweet がなければ、その点を NOISE とみなす (図 1-(c))。
4. 複数のクラスタが同一の Tweet を 1 つ以上共有するならば、それらのクラスタを結合する (図 1-(d))。
5. 2~4 を繰り返して終了した後、クラスタを形成する点を高頻度滞在地とし、NOISE となった点を低頻度滞在地



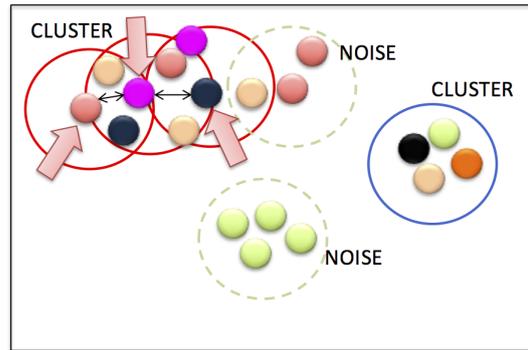
(a) 1 ユーザの全 Tweet の位置を抽出 (色の違いは投稿日の違いを示す).



(b) Tweet を 1 つ選び, その地点を中心とした円を描く, その円内に 4 日以上 の Tweet があればそれらを同じクラスタとする.



(c) 選択した Tweet の地点を中心とした円内に 4 日以上 の Tweet がなければ, その Tweet を NOISE とみなす.



(d) 2 つ以上のクラスタが互いに同じ Tweet を 1 つ以上共通に持つ場合, それらのクラスタを結合する.

図 1: DBSCAN の改良

とする。図 1-(d) にて, 高頻度滞在地は左上のクラスタと最も右のクラスタである。低頻度滞在地はそれ以外の NOISE に分類されたものである。

ここで, 最も多くの日数の Tweet を持つクラスタ (図 1-(d) の左上のクラスタ) の重心をそのユーザの居住地とし, 低頻度移動地を旅行・外出目的の訪問地とする。2014 年 8 月の日本での Tweet 件数は 13,299,546 件であった。これに対し, DBSCAN を実施し, 居住地および 1 つ以上の訪問地を持つユーザを抽出したところ, そのユーザ数は 107,418 人であった。

各ユーザの訪問地が得られた後に, 全ユーザについて訪問地が密集する地点を観光地として抽出する。Crandall ら [Crandall 09] は Flickr のジオタグ付き写真データの密集する地点を Mean Shift Clustering によってクラスタリングし, その中心地をランドマークとして抽出している。同様の方法で, 訪問地の密集する場所を集約する。今回のデータに対してクラスタリングした結果, 訪問地は 2336 の地点に集約された。

3. 各地点の特徴決定

訪問地を集約することで得られた観光地が周辺にどのような施設を持つかによって, その地域を特徴づける。Lee ら [Lee 13] は位置情報ベースのソーシャル・メディアである Yahoo! ロコによって, 各地域の特徴付けを行っている。本研究では, Foursquare によって各地域の特徴付けを行う。Foursquare はユーザが自分の訪問した場所 (Foursquare ではこれらをまとめて Venue と呼ぶ) について評価をするソーシャル・メディアである。各 Venue はレストランやホテルや史跡などといった施設のカテゴリを持つ。本研究では, Airport, Beach, Event,

Food, Historic Site, Hotel, Museum, Nightlife Spot, Outdoors & Recreation, Rest Area, Shop & Service, Stadium, Theme Park の 13 カテゴリを用いる。

図 2 は特徴付けの方法を示すものである。緑色の丸い点は Twitter のデータから得られた観光地である。各観光地の周辺 5km について, 各カテゴリごとの評価件数を総和する。この図では左の地点は史跡のポイントが 9, 右は 4 となる。さらにこれらの対数をとった上で, 各カテゴリの最大値で正規化することで特徴付けを行う。図 3 はこの手法で, 羽田空港周辺の特徴を表したものである。図の上の 5 つの単語はこの地域で投稿された Tweet のテキスト情報をつなぎ合わせて文書を作成した後に, 他の地域の文書と比較し, 最もこの地域を特徴づける単語を TF-IDF によって得たものである。この地域においては, 空港 (Airport) の特徴が最も高く, 確かに地域の特徴を捉えていることがわかる。

4. 決定木の作成と比較

どのような特徴の組み合わせがあれば観光客が集まるかを分析する。本論文では, 史跡 (Historic Site) の特徴値が最も高い上位 5% の地点 (133 地点) を取り出して, それらの地点における各特徴値が外国人観光客数と国内観光客数に与える影響を検討する。ここで Breiman ら [Breiman 84] によって提案された決定木の手法を用いる。決定木とは, 目的変数と説明変数の関係を表す際に, 説明変数のどのような組み合わせが揃うと目的変数が決定されるかを木構造で表すものである。これにより, 人間にとっても目的変数の決定条件が理解可能になる。ここでは, 目的変数を外国人観光客数と国内観光客数の 2 つの決定木を作り, とともに説明変数を各カテゴリの特徴値とする。

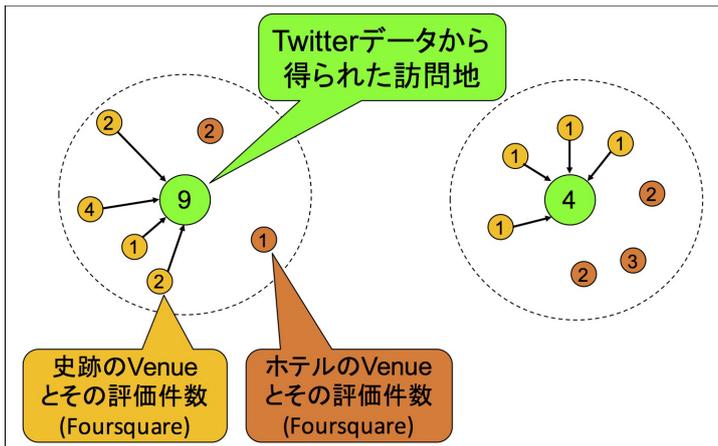


図 2: 各地域の特徴付与の方法

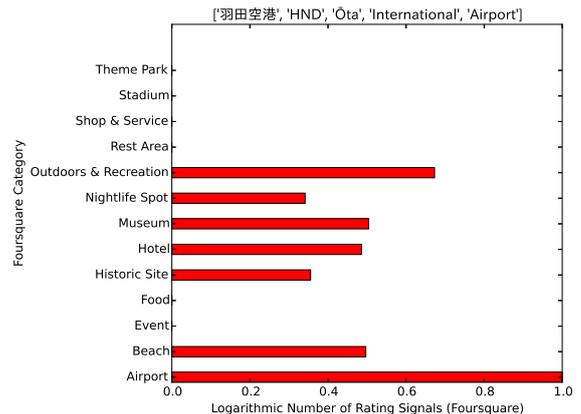
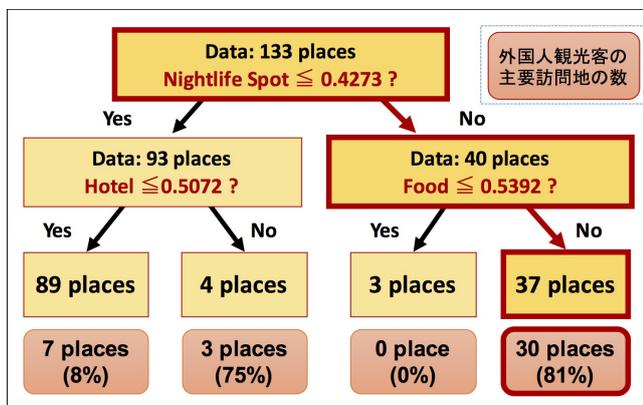
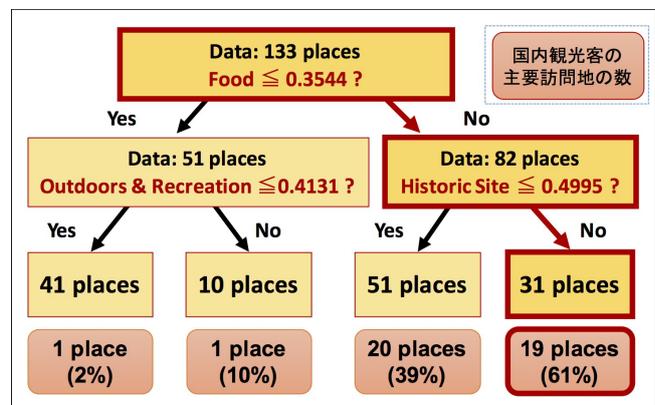


図 3: 周辺地域の特徴決定の例 (羽田空港)



(a) 外国人観光客



(b) 国内観光客

図 4: 史跡における観光客数の決定木: 外国人旅行者はナイトスポット (バー・ナイマーケット・クラブなど) を重視する。国内観光客においてはナイトスポットは重要でない。

外国人観光客のユーザ抽出には、この時期 (2014 年 8 月) に投稿された全世界のジオタグ付き Tweet データを利用した。ジオタグ付き Tweet データは、投稿時の緯度・経度以外に、投稿場所の国名を保持している。このため、日本において Tweet を投稿したユーザがその前後で日本以外の同じ国において Tweet を投稿しているならば、そのユーザは日本に外国から訪れた観光客とみなすことができる。このようにして外国人観光客を抽出する。

図 4 が求めた決定木である。図 4(a) は外国人旅行者数の決定木である。第 1 階層にて、133 地点がナイトスポット (Nightlife Spot) の値が 0.4273 以下か否かで、第 2 階層で左に移るか、右に移るかが分かれる。133 地点のうち 40 地点が 0.4273 より大きいため第 2 階層で右側に移る。ここでさらにレストラン (Food) の値が 0.5392 以下か否かで第 3 階層で左に移るか右に移るかが決まる。40 地点のうち 37 地点が第 3 階層で右側に移る。本論文では第 3 階層で終了するものとする。この図の第 3 階層のさらに下にある説明は、外国人観光客の主要訪問地 (全 2336 地域のうち、外国人観光客数が Top5% の地点と定義する) がそれぞれの分類にどの程度含まれているかを示すものである。第 3 階層の最右端の分類においては、37 地点のうち 81% である 30 地点が外国人観光客の主要訪問地であることを示す。この図から、最右端の分類に最も外国人観光客の主要訪問地が含まれており、この分類に入るためにはナイ

トスポット (Nightlife Spot) とレストラン (Food) の特徴量が高いことが条件であることがわかる。このため、外国人観光客はナイトスポットとレストランを重視しているのがわかる。同様に図 4(b) から国内観光客はレストランと史跡そのものを重視していることがわかる。このため、外国人観光客と国内観光客の大きな違いは、外国人観光客が史跡の周囲にナイトスポットがあることを望むのに対して、国内観光客は史跡の周囲にナイトスポットを望まず、むしろ史跡そのものの良さを求めるということである。したがって、日本に居住する人々が好んで行く史跡について、周囲にナイトスポットがなければ、外国人観光客はその史跡を敬遠してしまう可能性があるといえる。

5. 周辺環境の改善が必要な観光地

図 5 は、外国人観光客の主要訪問地でないが、国内観光客の主要訪問地であった場所を 4 つ取り上げたものである。これらの地域においては、レストラン (Food) の値は、図 4 より外国人観光客の基準も国内観光客の基準も満たしている。また、史跡 (Historic Site) の値も国内観光客の基準を満たしている。しかし、ナイトスポット (Nightlife Spot) の値が外国人観光客の基準を満たしていない。そのため、これらの地域においてはナイトスポットの改善が外国人観光客を引きつけるために必要であると考えられる。

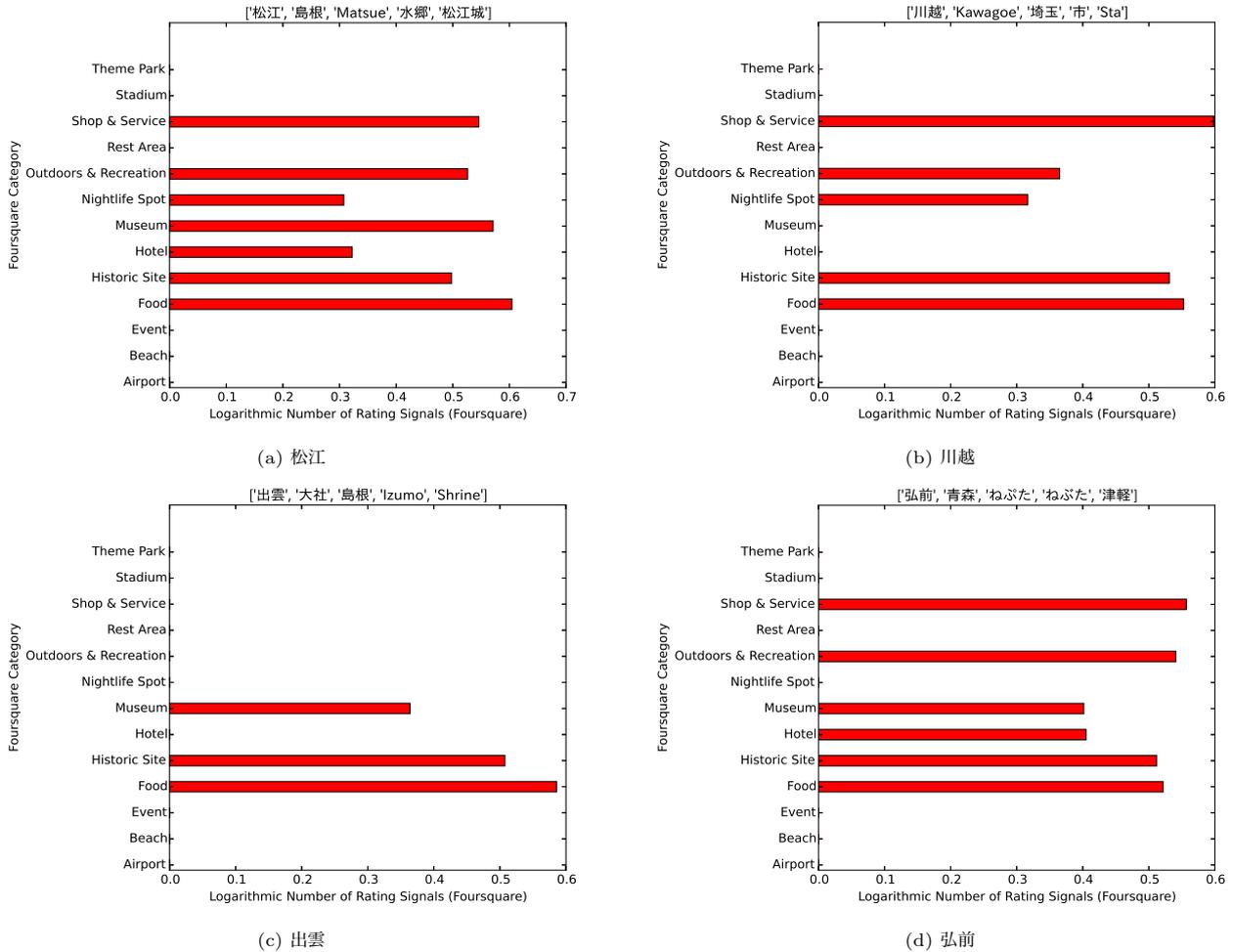


図 5: ナイトスポットを増やすことで外国人観光客数を増やすことができると考えられる観光地

6. おわりに

本研究では、ジオタグ付き Tweet データと Foursquare の Venue 情報をもとに外国人観光客と国内観光客が史跡の周辺に求める条件の違いを明らかにし、いかにすれば外国人観光客があまり訪れない観光地に外国人観光客を引きつけることができるかを示した。日本においてはとりわけナイトスポットが外国人観光客にとって重要であり、この点を改善することで、外国人観光客数を改善する可能性があることを示した。

今後は、季節ごとの違いや、出身国ごとの違いを明らかにすることで、きめ細かな分析を行うことが必要である。

参考文献

- [Breiman 84] Breiman, L., Friedman, J., Stone, C. J. and Olshen, R.: *Classification and Regression Trees*, Chapman and Hall/CRC (1984)
- [Crandall 09] Crandall, D. J., Backstrom, L., Huttenlocher, D. and Kleinberg, J.: Mapping the World's Photos. In *Proceedings of the 18th International Conference on World Wide Web*, pp. 761-770 (2009)
- [Ester 96] Ester, M., Kriegel, H.-P., Sander, J. and Xu, X.: A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp. 226-231 (1996)
- [Georgiev 14] Georgiev, P., Noulas, A. and Mascolo, C.: Where Businesses Thrive: Predicting the Impact of the Olympic Games on Local Retailers through Location-based Services Data. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (2014)
- [Lee 13] Lee, R., Wakamiya, S and Sumiya, K.: Urban Area Characterization Based on Crowd Behavioral Lifelogs over Twitter. *Personal and Ubiquitous Computing*, Vol. 17, pp. 605-620 (2013)
- [Paldino 15] Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C. and González, M. C.: Urban Magnetism through the Lens of Geo-tagged Photography. *EPJ Data Science*, Vol. 4 (2015)
- [Vu 15] Vu, H. Q., Li, G., Law, R. and Ye, B. H.: Exploring the Travel Behaviors of Inbound Tourists to Hong Kong Using Geotagged Photos. *Tourism Management*, Vol. 46, pp. 222-232 (2015)