

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

A Comparison of POIs Obtained from Multiple Social Medias with Location Information

大久保 航太郎† 望月 崇廣† 渡邊 貴之‡
Kotaro Okubo Takahiro Mochizuki Takayuki Watanabe

1. はじめに

近年, Flickr[1]などの位置情報付き画像ファイルの投稿が可能な写真共有ソーシャルメディアが普及している. また, これらの投稿から得られる位置情報をデータ源とした観光行動の分析の研究に注目が集まっている[2]-[5]. 投稿された写真群に紐付いた位置(経度・緯度)情報を基に, 多数のユーザが興味を持って撮影を行った場所である POI (Point of Interest) や, POI 間の回遊を表す POI ネットワークを抽出することができる. ここで, 写真共有ソーシャルメディアへの投稿を行ったユーザの大多数が観光客であると仮定すれば, これらの分析によって観光客の観光スポット間の周遊・回遊行動のモデル化が可能になると考えられる. また, 作成されたモデルを用いることで, 観光客の現在位置に即した情報の提供や観光スポットの推薦システムの実現が期待できる.

一方, 全ての観光客が, 訪れた全ての観光スポットで撮影を行い, 写真共有ソーシャルメディアに逐一投稿を行っているとは考え難い. さらに, 投稿には地元在住者などが自宅等で撮影した観光以外の写真も含まれている. そのため, 写真共有ソーシャルメディアから得られた POI 等の抽出結果が, 現実の観光客の周遊・回遊行動を精度よく反映しているかは不明であった. そこで, 我々は文献[5]において, 静岡県伊豆半島の南東部地域を対象とした 100 名の観光客に対するプローブパーソン調査によって収集した GPS ロガーデータと, 写真共有ソーシャルメディアである Flickr から得られた位置情報付き写真群を基に POI を抽出し, それらについて比較を行った. 結果から, Flickr から得られた POI および POI ネットワークは, プローブパーソン調査によって得られたものと類似した特徴を示すとの結論を得た.

しかし, 上記の結論は, Flickr 以外のその他の写真共有ソーシャルメディアにおいても合致するかは不明である. 例えば, 文献[6]では Flickr から得られた写真群から観光情報源として用いるには不適切な写真を人手で選別したところ, サンプル数 13,534 枚のうち非不要写真が 12,091 枚(89.3%)であったと報告している. すなわち, Flickr では観光情報源として不適切な写真は約 1 割に留まっており, 元から観光行動分析のデータ源として適したソーシャルメディアであったと考えられる. Flickr はスマートフォンが普及する以前の 2004 年に開始されたサービスであり, 当初から一眼レフなどの高画質デジタルカメラで撮影した写真の保管・共有のために利用されてきた[6]. 従って, 日常生活を撮影した写真を投稿するというよりは, 観光スポット等での非日常的な風景を投稿するために利用されてきたと言える.

本研究では, Flickr と Flickr 以外の写真共有ソーシャルメディアをデータ源として POI を抽出し比較を行う. 具体的には, Flickr とは特性の大きく異なると考えられる写真共有

ソーシャルメディアである Instagram[6]との比較を行う. Instagram は, スマートフォンの普及に合わせて爆発的に登録ユーザ数を獲得しているサービスであり, 観光情報分析のデータ源としては Flickr に比較して量的な側面から優位性を持つと言える. 一方で, 投稿可能な画像の縦横比が正方形に制限されるなど, 撮影した写真の加工が前提のサービスである. 文献[6]では, Flickr において不要写真と判定された画像の 68.9%が, Instagram との連携によって投稿されたものであったと報告しており, Instagram では私的な写真を共有する際にアプリを用いている傾向があると論じている. 本研究では, 観光行動分析にとって相対的に条件の厳しい Instagram から POI を抽出し, Flickr と比較することで, 条件の優劣によらないロバストな観光行動分析への活用可能性について議論する.

2. 関連研究

Flickr 以外の写真共有ソーシャルメディアを用いた観光分析研究としては, Panoramio を用いた文献[7]がある. Panoramio は Google Earth/Maps への掲載を前提としており, 風景写真であることや正確な位置情報が付けられているかなど掲載には審査が必要となる. 従って, Flickr と同様に観光行動分析のデータ源として適したソーシャルメディアであると考えられる.

次に, 文献[6]では, Flickr に投稿された写真から, 観光ポテンシャルマップ作成にとって適切な写真を自動的に選別する研究が行われている. 投稿に紐付いたメタデータの形式から 8 個のフィルタを作成し, 不要写真の自動除去を行っている. フィルタの 1 つとして Instagram または Instagramapp というタグを有する写真をすべて排除するという Instagram フィルタを設けており, Instagram が観光行動分析にとって条件の厳しいソーシャルメディアであることがわかる.

3. データセット

本研究では, Flickr と Instagram から得られた位置情報付き画像ファイルの投稿をデータ源として用いる. また, データの収集範囲は, 文献[5]と同様に静岡県伊豆半島とした.

図 1 は, Flickr 及び Instagram から得られた, 伊豆半島の位置情報が付与された写真を投稿したユーザ数を日付ごとにグラフ化したものである. 2015 年 3 月のユーザ数では Flickr が 50 人に対して, Instagram は 3,670 人と 70 倍以上の差が得られた. 従って, Instagram は観光情報分析のデータ源として Flickr に対して量的な側面から優位性を持つと言える.

次に, 伊豆半島全域を対象として得られた投稿データから, 図 2 で示された河津・下田エリアで 1 度でも投稿を行ったユーザのその当日の投稿のみを抽出し, 同一ユーザによる同一の位置情報を持つ投稿は最初の 1 件のみ残し, 重複した位置情報を持つものについては除外した. また, 同じユーザの投稿であっても日付が異なる投稿については別

† 静岡県立大学経営情報学部, School of Management and Information, University of Shizuoka

‡ 静岡県立大学 ICT イノベーション研究センター, Research Center for ICT Innovation, University of Shizuoka

ユーザの投稿とした。位置情報ポイントをマップ上にプロットしたものを図2に、ユーザ数等の比較を表1に示す。

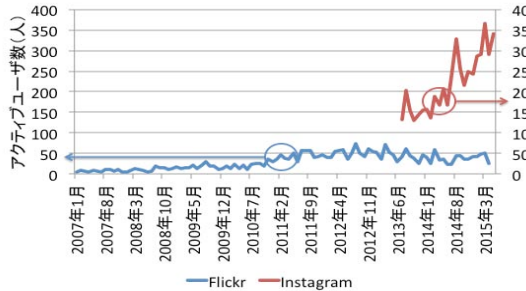


図1. 伊豆半島を対象とした投稿者数の推移

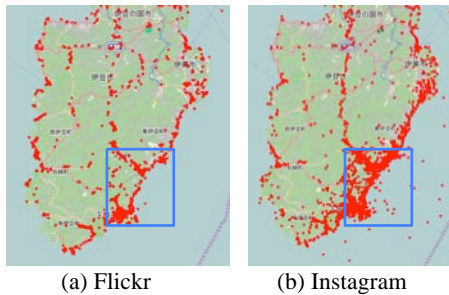


図2. 位置情報ポイントの比較

表1. 河津・下田エリアを対象としたデータセットの比較

項目	Flickr	Instagram
総ユーザ数	656人	9,209人
移動したユーザ数	352人	2,682人
写真総数	10,425枚	16,356枚
総ポイント数	5,927地点	14,770地点
平均移動距離	14.61km	4.63km

4. POIの抽出結果とその比較

取得したデータを用いて、Mean-shift法[8]によりPOIを抽出した。結果から、特に伊豆南部の主要な観光スポット集積エリアである下田エリアについて注目し比較を行う。

下田エリアのPOIの中で、位置情報ポイントをより多く含むPOI上位10位までを図3に示す。ただし、Instagramについては後述するように多くの不要写真の位置情報ポイントが含まれるため、具体的なポイントについては非表示としている。図中の地点名は、Foursquare APIを用いてPOIの経度・緯度を指定して得られた地点名の候補から、著者が選択したものである。ただし、適切な候補が得られなかったPOIには地点名を付していない。結果から、Instagramでは「寝姿山」「下田公園」「下田海中水族館」などの観光スポットがPOIとして抽出されていない。また、InstagramでPOIの抽出に使用された写真を確認すると、個人宅で撮影された写真など、観光写真ではない日常風景を撮影した写真が数多く存在した。

そこで、Instagramから得られた16,356枚の写真群から、観光情報源として用いるには不適切な写真を人手で選別した。選別の条件は文献[5]に準じたが、観光情報源に適しているが位置情報が不正確な写真も不要写真として判定した。また、判断困難な写真は不要写真に算入している。選別の結果、非不要写真は9,295枚(56.5%)となり、約半数は観光情報源として不適切な写真であることが確認された。不要写真を除去して抽出しなおした上位POIを図4に示す。



図3. 下田エリアでの上位10位までのPOI



図4. 不要写真を除去したInstagramにおける上位POI

5. まとめ

本研究では、Flickr及びInstagramから得られた静岡県伊豆半島におけるPOIについて比較した。結果から、写真共有ソーシャルメディアでは、不要写真の含まれる割合がサイト毎に異なっており、特にInstagramのように条件の厳しいサイトではデータの選別が必須である。一方、Flickrで有効とされるExif情報を用いた不要写真の自動選別は、InstagramではExif情報が取り除かれて掲載されるため使用できない。今後、メタデータが限られた条件での自動選別手法について検討したい。

謝辞

本研究は、総務省SCOPE(No.142306004)の助成を受けた。ここに深謝する。

参考文献

- [1] Flickr, <https://www.flickr.com/>
- [2] D.Crandall, et al., "Mapping the world's photos," Proc. 18th International World Wide Web Conference, no.10, pp.761-770, 2011.
- [3] 高木ら, "ジオタグ付き写真を用いたホットスポットを訪問する前後のユーザ行動の分析", 第7回データ工学と情報マネジメントに関するフォーラム(DEIM Forum 2015), 2015年3月.
- [4] 望月ら, "観光行動モデリングのためのプローブパーソンデータと位置情報付きSNSから得られたPOIの比較", マルチメディア, 分散, 協調とモバイル(DICOMO2015)シンポジウム, 2015年7月.
- [5] 倉田, "観光ポテンシャルマップ作成のための写真共有サイト投稿写真の自動選別", 観光情報学会第6回研究発表大会, 2012.
- [6] Instagram, <https://instagram.com/>
- [7] 長尾, "CGMをベースとした観光情報提供方法に関する考察", 観光情報学会第9回全国大会発表概要集, pp.26-27, 2012.
- [8] D.Comaniciu, and P.Meer, "Mean Shift: A robust approach toward feature space analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, pp. 603-619, 2002.