

観光ポテンシャルマップの信頼性向上に向けて  
—ソースとなる投稿写真データの自動選別ルールの構築—  
倉田陽平

**Toward More Reliable Potential-of-Interest Maps: Development of Rules for  
Automated Selection of Shared Photos for a Data Source**

**Yohei KURATA**

**Abstract:** Our previous work demonstrated that making use of a large number of geo-tagged photos shared on the Web, we can make *potential-of-interest maps*, which visualize the distribution of attractive locations for tourists. In order to increase the reliability of such potential-of-interest maps, this paper develops a decision tree with which we can remove inappropriate photos from the maps' data source.

**Keywords:** Flickr, 写真共有サイト (photo-sharing sites), 決定木 (decision tree), 観光ポテンシャル (potential-of-interest), ヴィジュアライゼーション (visualization)

1. はじめに

旅行者の多くは何かに興味を持った地点で写真撮影を行う。それゆえ、不特定多数の旅行者の撮影位置データを集約すれば、観光地内のどの地点がどの程度魅力的な場所なのかを見出せる可能性がある(倉田 2011)。一方、Flickrに代表される写真共有サイト上には既に膨大な位置情報付き写真が蓄積されている。そこで筆者は、Flickrに投稿された位置情報付き写真データから、非居住者による人気撮影箇所をカーネル密度関(ヒートマップ)として可視化し、これを旅行者案内に利用することを提案した(倉田 2011)。このようにして作られた「観光ポテンシャルマップ」は、観光地内の有名無名な見所や旅行者動線を視覚的に浮き彫りにする(図-1)。さらに長尾ら(2012)は「多数の人々が写真撮影した箇所」と「少数の人々が大量に写真撮影している箇所」とを描き分けることで、一般向けの見所と専門性の高い見所を可視化できる可能性を示した。

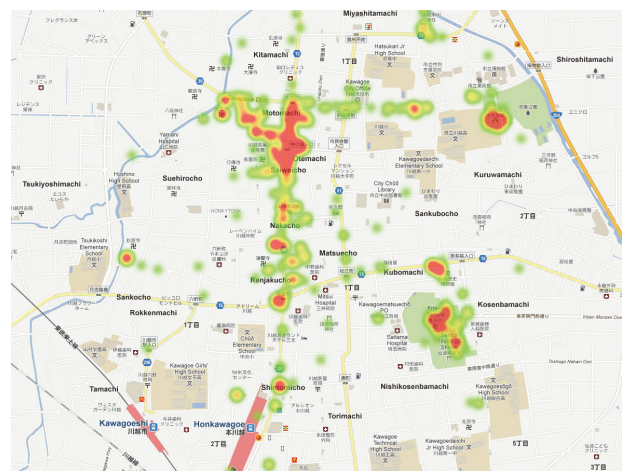


図-1 観光ポテンシャルマップの例 (川越)

観光ポテンシャルマップを作成する上で課題となるのは、データソースとなる写真の妥当性である。実はFlickr上で一般公開設定がなされている写真の中にも、自宅や職場で撮影した私的な写真など、観光ポテンシャルを評価するには不適切な写真が1割程度含まれている。そこで先行研究(倉田 2012)では、写真の属性データを利用し、不要な写真を排除するフィルタを8つ提案したが、

各フィルタを単体もしくは単純に組み合わせた場合の評価しか行わなかった。これをふまえ、本研究では、不要写真を判別するためのルール系列（決定木）を構築する。あわせてルール構築のための学習データを位置情報付きのものに写真に限定し、その量も増加させることで、信頼できるルール系列の構築に取り組む。

本論文の構成は以下の通りである。まず2章でサンプル写真データの入手手順について述べる。次に3章で、不要写真と有効写真の性質を比較する。この結果をふまえ、4章では判別ルールを構築し、その評価を行う。最後に5章で結論および今後の展望について述べる。

## 2. サンプル写真の抽出

まず、Flickr API を利用し、2012年9月および2013年5月中に札幌・仙台・横浜・名古屋・京都・大阪・神戸・広島・福岡の各都市中心部（市役所所在地から半径5km圏内）で撮影された写真、ならびに各写真の撮影者が同日中に撮影した写真をFlickr上の公開写真群からすべて抽出した。得られた計28361枚の写真それぞれについて、それが同一撮影者によってその日撮影された写真群のうち何枚中何枚目かを記録した。また、撮影者の居住地が登録されているものは、それについても記録した。

次に、この28361枚の写真から、位置情報が付与されていないものを除去した。つづいて、残った19019枚分の写真に対し、それぞれ画像と属性データを人力で確認し、不要写真1113枚(5.9%)、有効写真17607枚(92.6%)、判断困難写真299枚(1.6%)へと振り分けた。先行研究で述べたとおり、旅行時に撮影した写真か否かを客観的に判別することは困難であるため、代わりに「ある地域において一般の人々がアクセスできる空間を撮影したものであれば、旅行先としてその地域を知るうえで何らかの手がかりとなる可能性がある」とし、それらを有効写真か否かの基準とした。言

い換えれば、不要写真は、私的性の高い空間（たとえば自宅や職場、ホテルの室内）で撮影された写真や、実空間との関連性に乏しい写真（たとえばテレビやPCの画面写真、イラスト）などが該当することになる。

## 3. 不要写真と有効写真との比較

先行研究（倉田 2012）にならい、2章のサンプル写真を利用して、不要写真と有効写真を判別するために有効な属性データを探った。

まず表-1に、利用可能な主な属性データと、今回のサンプルにおけるそれぞれの記録率を示す。不要写真はタイトル、タグ、撮影者の居住地、ファイルソース、被写体距離レンジの記録率が相対的に低い。タイトル、タグの記録率が低いのは、私的な写真では投稿者がそれらを省略する傾向があるためではないかと推察される。

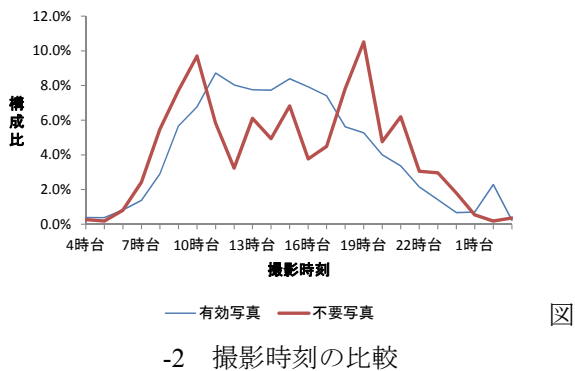
表-1 主な属性データの記録率

属性値	不要写真	有効写真	
Flickr側の メタデータ	タイトル	83.0%	92.2%
	タグ	51.1%	61.0%
	撮影時刻	100.0%	100.0%
	投稿者の居住地	46.3%	60.8%
EXIFデータ	ファイルソース	18.9%	33.2%
	機器メーカー	73.5%	75.4%
	機器名	73.0%	75.4%
	レンズf値	69.9%	73.7%
	焦点距離	70.1%	73.9%
	ISO感度	70.2%	74.0%
	フラッシュ	70.0%	73.6%
	露出モード	68.6%	73.4%
	露出プログラム	68.4%	66.2%
	ホワイトバランス	68.6%	73.4%
	被写体距離レンジ	10.3%	16.9%
被写体距離	1.3%	0.5%	

次に、表-1の属性データのそれぞれについて、有効写真と不要写真との比較を行った。以下では、顕著な差が見られた項目について述べる。

## 撮影時刻

有効写真は日中にピークがある一方で、不要写真は朝・晩にピークがあることがわかった(図-2)。この結果は、朝・晩は自宅や宿にて食事したりくつろいだりしている間に撮影が行われるためだと推察される。



## 居住地・撮影地間の距離

撮影者居住地のわかる写真について、居住地・撮影地間の距離を分析したところ、不要写真の場合は半数以上が10km圏内居住者によって撮影されたものだった(図-3)。これは、地元においては日常の私的な写真を投稿する傾向が高まるためではないかと推察される。逆に500km以上遠方在住の人々の写真は有効写真の割合が高かった。

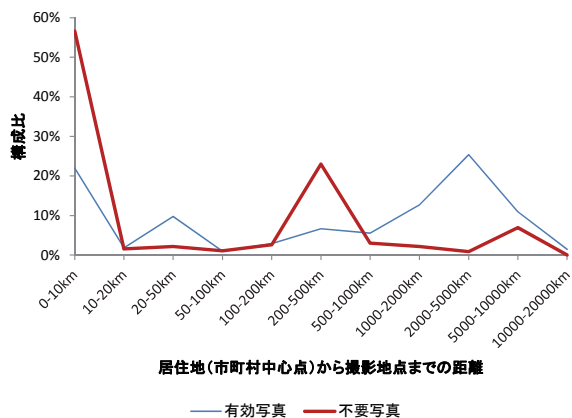


図-3 居住地・撮影地間の距離の比較

## 機器メーカー

機器メーカー名が記録されている写真14203枚について分析したところ、Apple社とPanasonic社の機器で撮影された写真において不要写真率

が著しく高かった(表-2)。Apple社については、iPhoneで撮影された日常生活の写真が多く含まれるためだと推察される。他社のスマートフォンによる写真でも同様のことが言えるかも知れないが、メーカー・機器が分散しているため、明らかな傾向は見出せなかった。また、Panasonic社のケースについては、偶然の結果の可能性が高い。

表-2 機器メーカー別の不要写真の含有状況

	写真枚数	不要写真枚数	不要写真率
Canon	4364	51	1.2%
Apple	3387	430	12.7%
NIKON	1919	109	5.7%
OLYMPUS	1214	41	3.4%
SONY	712	4	0.6%
FUJIFILM	688	1	0.1%
Panasonic	422	65	15.4%
その他	1497	117	7.8%
総計	14203	818	5.8%

## 不要写真発生タイミング

1日10枚以上撮影した人々の写真のうち、不要写真がどの位置で発生するのか分析したところ、末尾10%で出現割合が高まることがわかった(図-4)。これは、撮影者が自宅やホテルに帰ってきてから、夕食やその日入手した品物を撮影するケースが多くあったためと推察される。

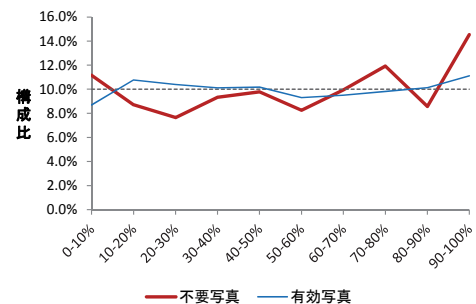


図-4 同一日に撮影した写真群の中での相対位置

## 4. 不要写真判別ルール構築

与えられた写真が観光ポテンシャルマップのソースとして適切か否かをその属性データから段階的に判別するために、決定木学習を利用する。

学習データには、正負事例のバランスをとるため、2章で抽出した不要写真全 1113 枚、および有効写真 17607 枚から無作為抽出した 1113 枚、計 2226 枚を利用した。説明変数としては、3章の議論をもとに、①タイトル有無、②タグ数、③ファイルソース有無、③撮影時刻（1時間おき 24 段階）、④居住地登録有無、⑤撮影地・居住地間までの距離（50km 圏内・500km 圏内・500km 圏外・データ無し）、⑥機器メーカー（Apple 社製・他社製・データ無し）、⑦被写体距離レンジ有無、⑧同一日写真群における相対位置（10%おき 10 段階）の 8 変数である。決定木算出には IBM SPSS Decision Trees 20 を利用した。学習アルゴリズムとして Exhaustive CHAID (Biggs *et al.* 1991) を使用し、木の最大深さは 5、各ノードの最小事例数 100 に設定した。この結果、図-5 に示す決定木が得られた。この決定木を見ると、不要写真・有効写真を区別する最大の手ごかりは居住地・撮影地間の距離である。そして、近距離居住者もしくは居住地未登録者の場合、Apple 製品によって撮影された写真かどうか、および撮影時刻が有力な手がかりになることがわかった。

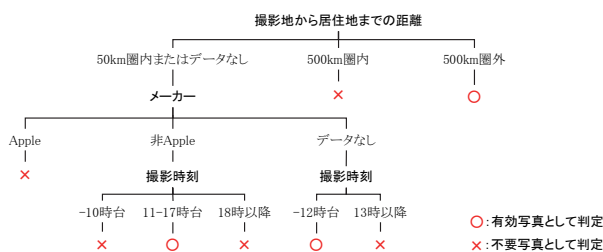


図-5 不要写真・有効写真の判定ルール

次に、得られた判定ルールを当初用意したすべてのサンプル写真（19019 枚）に対し適用したところ、不要写真のうち 86.3%が除去される一方で、有効写真のうち 49.1%が残存した。この結果、有効写真の割合は 92.6%から 97.5%へと向上、不要写真の割合は 5.9%から 1.7%へと減少し、判定ルールの有効性が確認された（表-3）。

表-3 写真判定ルール適用後の割合比較

	適用前	適用後
有効写真	92.6%	97.5%
不要写真	5.9%	1.7%
判定困難写真	1.6%	0.8%

## 5. おわりに

本論文では、Flickr 上の写真に対し、観光ポテンシャルを評価する上で有効か否かを判定するためのルールを導出した。この判定ルールを Flickr 上の膨大な写真群に適用することで、それらをもとに、より質の高い観光ポテンシャルマップを構築できる。観光ポテンシャルマップは、投稿写真データが十分にある限り、世界中どこでも作成することができるという点で、非常に可能性があるものである。そこで今後は、見せ方の改良を行いながら、観光における実用化と普及に取り組んでいきたい。また、本研究の応用として、Flickr のキーワード検索機能を利用し、テーマ別の観光ポテンシャルマップ作成についても今後取り組んでいきたい。

## 謝辞

本研究には科学的研究費補助金（課題番号：23701030 研究課題名：旅行者の写真撮影位置情報を利用した観光ポテンシャルマップの構築と観光案内への応用）を一部利用した。

## 参考文献

- 倉田陽平(2012):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回研究発表会
- 倉田陽平(2011) 観光ポテンシャルの可視化によるスマートフォン向けのシンプルなお観光情報サービス. 地理情報システム学会講演論文集 20, CD-ROM.
- 長尾 光悦(2012):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, 26-27.
- Biggs, D., B. De Ville, and E. Suen (1991). A method of choosing multi-way partitions for classification and decision trees. Journal of Applied Statistics, 18(1), 49-62.