

平成 26 年度修士論文

写真共有サイトの位置情報付き写真データを利用した
テーマ別観光マップの作成

Creation of Various Thematic Tourist Maps
Based on Geotagged Photos in a Photo-Sharing Site

首都大学東京大学院
都市環境科学研究科
観光科学域

13842404 真田風
指導教員 倉田陽平

目次

I. はじめに	1
II. 既存研究	2
2.1 位置情報付き写真への注目	2
2.2 位置情報付き写真を用いた観光情報研究	3
2.3 観光ポテンシャルマップ研究	4
III. 研究目的	6
IV. 調査 1 データの抽出と手動判別	7
4.1 方法	7
4.2 結果と考察	14
4.3 調査 1 のまとめ	32
V. 調査 2 自動選別ルール構築	33
5.1 方法	33
5.2 結果と考察	36
5.3 調査 2 のまとめ	50
VI. まとめと課題	51

図表目次

図 1	京都における観光ポテンシャルマップの例（倉田 2012 より）	5
図 2	夜（狭い定義）のマップ	20
図 3	夜（広い定義）のマップ	20
図 4	夜間（狭い定義）のマップ	20
図 5	夜間（広い定義）のマップ	20
図 6	雨（狭い定義）のマップ	22
図 7	雨（広い定義）のマップ	22
図 8	雨天時（狭い定義）のマップ	22
図 9	雨天時（広い定義）のマップ	22
図 10	雪（狭い定義）のマップ	23
図 11	雪（広い定義）のマップ	23
図 12	夕焼けのマップ	23
図 13	朝焼けのマップ	23
図 14	虹のマップ	23
図 15	猫のマップ	24
図 16	蝶のマップ	24
図 17	猫のマップ（秋葉原駅周辺）	24
図 18	桜のマップ	25
図 19	梅のマップ	25
図 20	紫陽花のマップ	25
図 21	銀杏のマップ	25
図 22	楓のマップ	25
図 23	桜 3 月のマップ	27
図 24	桜 4 月のマップ	27
図 25	桜 3,4 月以外のマップ	27
図 26	そばのマップ	28
図 27	寿司のマップ	28
図 28	ラーメンのマップ	28

図 29	着物のマップ	28
図 30	ラーメンのマップ (写真表示版)	29
図 31	コスプレのマップ	30
図 31	神輿のマップ	30
図 33	クリスマスツリーのマップ	30
図 34	富士山のマップ	31
図 35	東京スカイツリーのマップ	31
図 36	撮影地から居住地までの距離 (夜 (狭い定義))	36
図 37	撮影時刻 (夜 (狭い定義))	36
図 38	撮影月 (夜 (狭い定義))	36
図 39	シャッタースピード (夜 (狭い定義))	36
図 40	夜 (狭い定義) 写真選別の決定木	37
図 41	撮影地から居住地までの距離 (夜 (広い定義))	38
図 42	撮影時刻 (夜 (広い定義))	38
図 43	撮影月 (夜 (広い定義))	38
図 44	シャッタースピード (夜 (広い定義))	38
図 45	夜 (広い定義) 写真選別の決定木	39
図 46	撮影地から居住地までの距離 (桜)	40
図 47	撮影時刻 (桜)	40
図 48	撮影月 (桜)	40
図 49	シャッタースピード (桜)	40
図 50	桜写真選別の決定木	41
図 51	撮影地から居住地までの距離 (猫)	42
図 52	撮影時刻 (猫)	42
図 53	撮影月 (猫)	42
図 54	シャッタースピード (猫)	42
図 55	猫写真選別の決定木	43
図 56	撮影地から居住地までの距離 (ラーメン)	44
図 57	撮影時刻 (ラーメン)	44
図 58	撮影月 (ラーメン)	44

図 59	シャッタースピード（ラーメン）	44
図 60	ラーメン写真選別の決定木	45
図 61	撮影地から居住地までの距離（東京スカイツリー）	46
図 62	撮影時刻（東京スカイツリー）	46
図 63	撮影月（東京スカイツリー）	46
図 64	シャッタースピード（東京スカイツリー）	46
図 65	撮影地から対象物までの距離（東京スカイツリー）	46
図 66	東京スカイツリー写真選別の決定木	47

表 1	各テーマの検索ワードと手動判別基準（キーワードによる抽出）	12
表 2	各テーマの手動判別基準（条件による抽出）	13
表 3	各テーマの手動判別結果（キーワード検索による抽出）	14
表 4	各テーマの手動判別結果（条件による抽出）	15
表 5	各テーマのジニ係数	18
表 6	決定木学習法に用いるサンプルデータ数	34
表 7	属性値の記録率（夜（狭い定義））	36
表 8	機器メーカーごとの不要写真率（夜（狭い定義））	36
表 9	属性値の記録率（夜（広い定義））	38
表 10	機器メーカーごとの不要写真率（夜（広い定義））	38
表 11	属性値の記録率（桜）	40
表 12	機器メーカーごとの不要写真率（桜）	40
表 13	属性値の記録率（猫）	42
表 14	機器メーカーごとの不要写真率（猫）	42
表 15	属性値の記録率（ラーメン）	44
表 16	機器メーカーごとの不要写真率（ラーメン）	44
表 17	属性値の記録率（東京スカイツリー）	46
表 18	機器メーカーごとの不要写真率（東京スカイツリー）	46
表 19	構築したルール、東京におけるデータへの適用結果	48
表 20	構築したルール、大阪におけるデータへの適用結果	48

要旨

近年観光に対するニーズは多種多様化している。しかし、雑誌などで提供されている観光情報は、多数の旅行者が望む一般的な観光情報であり、多様化する観光ニーズに対応できているとは言い難い。これに対し、CGM（Consumer Generated Media）におけるデータを用いることで、効果的な観光情報が獲得できると考えられる。中でも、Flickr やフォト蔵をはじめとする写真共有サイトの利用拡大や、GPS センサ搭載携帯デバイスの普及により、位置情報の付いた写真が Web 上に大量にアップロードされており、それらの写真データを活用した観光情報の研究が盛んに行われている。例えば、このような研究の例としては Web 上の大量の写真を画像分類して提示する観光マップの提案や、写真撮影の位置軌跡を利用した旅程推薦システムの開発、写真群の位置情報を元に人気撮影箇所をヒートマップとして可視化する観光ポテンシャルマップの作成等がある。しかしいずれの研究も、大量の写真データを用いて、大勢の人が注目した場所を見出し、観光情報として提供することを目的としているために、近年の多様化する観光ニーズへの対応は難しい。

一方で、一連の観光ポテンシャルマップ研究の中で「写真のタイトルやタグの中身に踏み込み写真を選別する、テーマ別のマップ作成の可能性」が指摘されている。そこで本研究では、観光ポテンシャルマップ研究を土台に、テーマ別観光マップの作成に取り組む。そして、低コストで有用な CGM データを用いた、テーマ別観光マップ作成の方法論を導き出すことを目的とする。

テーマ別観光マップを作成するにあたり、まずはどのようなテーマが実際に地図化に適しているのかを評価する必要がある。さらに、抽出データ数の多いテーマに対しては、先行研究同様、テーマに適した有効な写真を自動的に選別するためのルール構築が必要である。これは、枚数が多すぎると人力での判別が追い付かず、今後の他地域や他期間への応用が難しいからである。以上をふまえ、調査は 2 段階で行った。

まず調査 1 では、地図化が期待できそうなテーマの候補を選定し、対応するキーワードによってサンプルとなる写真を抽出し、さらに目視によって有効写真と不要写真の判別を行った。対象地域は東京駅を中心に東西 10km、南北

10km の範囲とし、2013 年の 1 年間に撮影され Flickr に投稿された写真を用いた。地図化するテーマとして選定したのは、夜、雨、猫、桜、ラーメン、コスプレ、東京スカイツリー、などをはじめとした計 23 テーマである。各テーマについて、有効写真枚数、有効写真率、有効写真の投稿者数と投稿写真枚数の偏り、等のデータを元に、テーマとしての地図化適性を考察した。

次に調査 2 では、調査 1 において有効写真枚数の多かった、夜、猫、桜、ラーメンの 4 テーマについて、決定木学習法を用いて、有効写真を自動選別するためのルールを構築した。説明変数として、Flickr に記録されている写真のメタデータ（写真のタイトル、タグ、撮影日時等）や個々の写真に記録されている Exif データ（撮影機材のメーカー、シャッタースピード等）を用いた。構築したルールは、他地域（大阪）で抽出されたデータに適用することで、その有効性を検証した。

以上の 2 つの調査により、①虹のようなテーマは、検索ワードに多義性があり、有効写真率が低くなるということや、②猫やコスプレなど、同一被写体が様々なポーズや表情を見せるようなテーマは、一部の熱狂的なファンにより同じような写真が複数枚投稿されるケースが見られ、投稿者ごとに投稿枚数に偏りが出る、というようなことがわかった。また、自動選別についてわかったことの例としては、①夜の有効写真を自動で選別するためには、シャッタースピードという、そのテーマの写真の特性に合った説明変数が有効で、他地域への応用も利きやすいということや、②一部の人による大量写真投稿がある猫などのテーマは、投稿者の情報に依存しない説明変数を用意することが、他地域にも応用できるルールを構築するためには必要である、ということが挙げられる。

本研究では地図化を行いたいテーマごとに、抽出されたデータを分析し考察することで地図化適性に影響する要因を整理し、データ数が多いテーマに対しては自動選別ルールの構築を行い、有効なルール構築に必要な説明変数の性質を明らかにすることで、テーマ別観光マップの作成手法の一部を示した。

Creation of Various Thematic Tourist Maps Based on Geotagged Photos in a Photo-Sharing Site

Abstract

Recently, consumers' needs for tourism become more and more diverse. However, ordinary tourist information, such as those in magazines, cannot support the diverse needs of tourists, because such information is usually general tourist information that a large number of tourists want. In contrast, it is expected that useful tourist information can be extracted from CGM data. In particular, many studies about the extraction of tourist information based on geotagged photos in photo-sharing sites have been conducted, thanks to the widespread use of the photo-sharing sites, such as Flickr and Photo-Zou, as well as the spread of the GPS-equipped mobile devices. Those studies include automatic generation of tourism maps which present the land characteristics by classifying a large number of photos on the web, development of an itinerary recommendation system using the trajectory of photo-shooting location data, creation of potential-of-interest maps which visualize popular photo-shooting locations as a heat map based on geotagged photo collections, and so forth. However, these studies hardly respond the recent diverse needs of tourists, because these studies focus on the places to which many people paid attention, using a large amount of photo data.

On the other hand, the possibility of generating thematic tourist maps making use of keyword search on photo titles and tags is pointed out in the studies of potential-of-interest maps. This study, therefore, attempts to create various thematic tourist maps based on the studies of potential-of-interest maps, in order to seek the methodology of generating thematic tourist maps from low-cost and useful CGM data.

At first, it is necessary to judge what kinds of themes are suitable for

creating thematic tourism maps. In addition, it is necessary to develop the rules for discerning effective photos automatically from a large number of the extracted photo data, because the distinction of effective photos by human power is costly when considering the creation of maps in other areas or periods. Thus, we conducted two investigations

First, we selected the candidates of map themes, extracted a collection of sample photos by keyword search for each theme, and distinguished effective photos in each collection by human eyes. We used the photos that have been posted on Flickr, taken in $10\text{km} \times 10\text{km}$ area around Tokyo Station in the year 2013. We selected 23 map themes, such as "night", "rain", "cat", "cherry blossom", "ramen", "cosplay", and "Tokyo Skytree". For each theme, we discussed the suitability for mapping, considering the total number of effective photo, ratio of effective photo, and deviation of the number of photos posted by users.

Then, we developed the rule for distinguishing effective photos automatically using decision trees for four themes with a large number of effective photos: "night", "cat", "cherry blossom", and "ramen". As explanatory variables, we used the metadata of photos recorded by Flickr (title, tags, shooting date and time, etc.), as well as Exif data recorded in the data of individual photos (manufacturer of photographic equipment, shutter speed, etc.) We examined the effectiveness of the developed rules by applying them to the data in other area (Osaka).

As a result of the first investigation, we revealed i) the themes expressed by polysemic words, such as "rainbow", lead to low effective photo, and ii) the themes with enthusiastic fans who post a large number of photos, such as "cat" and "cosplay", lead to the deviation in the number of photos posted by users. Then as a result of the second investigation, we revealed i) in order to discern automatically the effective photos of "night", the explanatory variables that match on the characteristics of the photos of the theme, such as "shutter speed", is effective, and ii) the themes with

enthusiastic fans, such as “cat”, requires the explanatory variables that do not depend on the characteristics of specific contributors.

In this way, this study clarified several factors that affect the suitability of mapping from the analysis of the extracted data for each theme. In addition, we developed several rules to discern effective photos automatically for several themes, and clarified the nature of the explanatory variables required for the development of effective rules.

I. はじめに

近年観光に対するニーズは多種多様化しているが、雑誌などで提供されている観光情報は、多数の旅行者が望む一般的な観光情報であり、多様化する観光ニーズに対応できているとは言い難い。これに対し、CGM（Consumer Generated Media）におけるデータを用いることで、効果的な観光情報が獲得できると考えられる（長尾 2012）。中でも、Flickr¹やフォト蔵²をはじめとする写真共有サイトの利用拡大や、GPS センサ搭載携帯デバイスの普及により、位置情報の付いた写真が Web 上に大量にアップロードされていることから、それらの写真データを用いた観光情報の研究に注目が集まっている。例えばこのような研究の例としては、Web 上の大量の写真を画像分類して提示する観光マップの提案（王ら 2011）や、写真撮影の位置軌跡を利用した旅程推薦システムの開発（奥山・柳井 2011）、写真群の位置情報を元に人気撮影箇所をヒートマップとして可視化する観光ポテンシャルマップの作成（倉田 2013 ほか）等がある。しかしいずれの研究も、大量の写真データを用いて、大勢の人が注目した場所を見出し、観光情報として提供することを目的としているために、近年の多様化する観光ニーズへの対応は難しい。

そこで注目したいのが、今後の展望として「テーマ別観光マップの作成」を挙げている、観光ポテンシャルマップの研究（倉田 2013 ほか）である。従来は「旅行中に撮影された写真」全てを用いることで人気写真撮影箇所を表現していたのに対して、使用する写真の抽出の際にキーワード検索を行うなどすることで、特定の観光資源に関するテーマ別観光マップを作成しようというアイデアが提起されている。そこで本研究では、この観光ポテンシャルマップ研究を土台に、テーマ別観光マップの作成に取り組みたい。

本論文は、以下のような構成になっている。第二章でまず、位置情報付き写真を用いた観光情報研究の既存研究と、本研究で注目する観光ポテンシャルマップ研究についてまとめる。第三章で、本研究の目的と調査の方針を示す。第四章、第五章で、写真共有サイトのデータ利用したテーマ別観光マップ作成のための調査について、2 段階にわけて方法、結果、考察を述べる。最後に第六章で、まとめと課題、今後の展望を述べる。

Ⅱ．既存研究

本章では、位置情報付き写真への注目、位置情報付き写真を用いた観光情報研究の紹介、本研究の土台となる観光ポテンシャルマップ研究（倉田 2013 ほか）についてまとめる。

2.1 位置情報付き写真への注目

倉島ら（2010）は、「写真を撮る」行為は実世界のある場所に対する投票とみなすことができ、写真撮影箇所のデータは訪れる価値のある場所を発見することが可能な質の高い情報源である、と述べている。倉田ら（2010）も、観光関心点の収集方法として、GPS 搭載型デジタルカメラを利用して、旅行者が写真撮影を行った地点を記録する方法の優位性を論じている。彼らが比較しているのは、①地図書込み方式、②現地手動記録方式、③GPS カメラ方式、④停留箇所検出方式、⑤テキストマイニング方式の 5 つの方法で、デバイス依存性、作業負担、位置精度、コミッションエラー、オミッションエラー、情報粒度一貫性という観点における問題点の少なさや、iPhone のような GPS・カメラ搭載型スマートフォンの普及が著しい昨今の時代背景を鑑みた結果である。

杉本（2011）は、観光行動研究のための写真撮影調査にデジタルカメラを使用することの意義や問題点について考察している。近年のカメラ市場のほとんどをデジタルカメラが占めており、デジタルの写真データには、Exif という撮影地点の位置データ、撮影した日時、使用したデジタルカメラの機種に関するデータ、撮影行動に関するデータなどの様々な情報が記録されている。そして、Flickr 等の写真コミュニティサイトの登場によって、デジタル写真を Web 上に公開するなど、デジタル写真を用いた楽しみ方が広まっている。Flickr では、写真を位置情報と結び付けて投稿することができる。このことから、観光行動においてもデジタルカメラを用いることが一般的になっており、研究においてデジタル写真を用いることの有用性、信頼性が杉本（2011）によって論じられている。

2.2 位置情報付き写真を用いた観光情報研究

杉本（2011）でも触れられていた通り、近年 Flickr やフォト蔵をはじめとする写真共有サイトの利用が拡大していることから、位置情報の付いた写真が大量に Web 上にアップロードされている。このようなデータは、ほぼコストなしに入手することができる。このことから、実際に観光情報の研究において、Web 上の位置情報付きの写真と地図を用いた研究が盛んである。そのような研究の例としては、観光資源を自動検出し観光マップを自動作成する研究（Chen, et al. 2009）や、写真を画像分類して土地特性を提示する観光マップ作成の研究（王ら 2011）、移動軌跡を参考に、旅程推薦を行う研究（De Choudhury, et al. 2010; 奥山ら 2011）や、先述の、倉田（2013 ほか）による写真群の位置情報をヒートマップとして可視化した、観光ポテンシャルマップの作成等がある。

しかしながらこれらの研究においては、大量の写真データを用いて、大勢の人が注目した場所を見出し、観光情報として提供することを目的としているため、抽出できる観光情報は一般的に高い注目を集めているものである可能性が高く、多様なニーズへ対応することは難しい。

2.3 観光ポテンシャルマップ研究

本研究で注目したいのは、先述の観光ポテンシャルマップ研究である。その理由を説明する前にまずは、一連の観光ポテンシャルマップ研究について整理しておく。

観光ポテンシャルマップは、新しい観光案内ツールとして倉田ら（2010）によって提案された。図 1 は、京都における観光ポテンシャルマップの例（倉田 2012 より）である。倉田（2010）は、観光関心点の収集方法として、位置情報付きの写真を用いることの優位性を論じ、便宜的に、写真共有サイト **Flickr** に投稿されている位置情報付きの写真を用いてそれをヒートマップとして可視化することで、横浜中心部の観光ポテンシャルマップのプロトタイプを作成した。それを元に、倉田（2011）ではスマートフォン向けの観光情報サービスとしてアプリを開発し、実証実験を行うことで、観光ポテンシャルマップへの一定のニーズを感じ取っている。これらの研究においては、写真共有サイト **Flickr** に投稿されている位置情報付きの写真を「旅行中に撮られた写真」とみなし、それを用いることで「他の旅行者が関心を持った場所」、すなわち観光ポテンシャルを可視化し、シンプルに見どころを示すことで、効果的な観光案内を目指している。しかし、**Flickr** に投稿されている写真が旅行者によって撮影されたものだという確証はないので、サービスの向上のためには、実際に旅行中に撮られた写真を用いることが必要だという課題が挙げられている。

そこで、倉田（2013）により、**Flickr** に投稿されている位置情報付き写真の中から、旅行中もしくはレジャー活動中に撮影された写真を自動的に選別するためのルールが構築された。方法としては、まずサンプル写真に対し人の手によって判別を行う。判別基準は「旅行先となりうる場所（公共空間）で撮られたか否か」である。そしてその判別結果を参考に、写真に付加されているメタデータ（写真のタイトル、タグ、撮影時刻等）や **Exif** データ（カメラのメーカー、被写体距離、フラッシュ利用の有無等）から、適切な説明変数を決め、決定木学習によって選別ルールを構築する。自動的に不要な写真を排除できるようになれば、**Web** 上の大量の写真を用いて、低コストであらゆる地域の観光ポテンシャルマップを作成できるようになる、というのが狙いである。そして、今後の展望として「写真のタイトルやタグの中身に踏み込み写真を選別する、

テーマ別のマップ作成の可能性」について述べている。

従来の観光ポテンシャルマップにおいては、「旅行中に撮影された写真」全てを用いることで人気写真撮影箇所を表現していた。これに対し、使用する写真の抽出の際にキーワード（タイトルやタグの情報）検索を行うなどすることで、特定の観光資源に関する観光マップが作成できるようになることが期待される。これが、本研究において観光ポテンシャルマップ研究に注目する理由である。

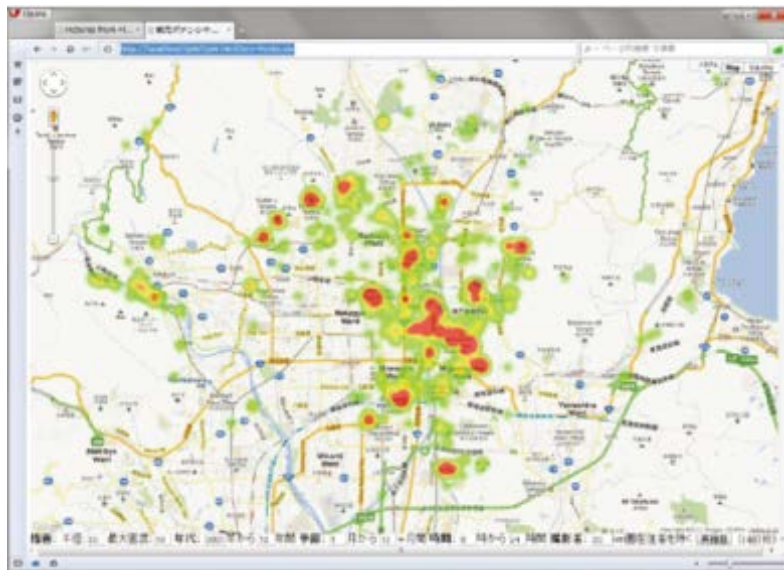


図 1 京都における観光ポテンシャルマップの例（倉田 2012 より）

Ⅲ. 研究目的

本研究では、観光ポテンシャルマップ研究を土台に、テーマ別観光マップの作成に取り組む。そして、低コストで有用な CGM データを用いた、テーマ別観光マップ作成の方法論を導き出すことを目的とする。

テーマ別観光マップを作成するにあたり、まずはどのようなテーマが実際に地図化に適しているのかを評価する必要がある。さらに、抽出データ数の多いテーマに対しては、先行研究同様、テーマに適した有効な写真を自動的に選別するためのルールの構築が必要である。これは、枚数が多すぎると人力での判別が追い付かず、今後の他地域や他期間への応用が難しいからである。

そのため、調査は 2 段階にわけて行う。1 段階目では、地図化したいテーマの選定と、そのテーマに沿った写真データの抽出と手動判別に関する調査を行い、2 段階目では地図化に必要なデータを自動で選別するための自動選別ルールの構築に関する調査を行う。

Ⅳ．調査 1 データの抽出と手動判別

4.1 調査方法

(1) 概要

第 1 段階では、地図化する特定のテーマを決め、サンプルとなる写真を抽出し、各テーマ別に人の手によって有効写真と不要写真の判別を行う。また、有効と判別された写真のみをマッピングして、テーマ別の、理想的な観光マップを表示し、それぞれの分布の特徴について考察する。具体的な分析の内容は以下の通りである。

①データの抽出数

各テーマにおける有効写真枚数や有効写真率に注目して、地図化適性を考察する。

②投稿者数と投稿枚数の偏り

各テーマにおける有効写真の投稿者数と各投稿者の投稿枚数のデータを元に、ジニ係数を求め、投稿枚数の偏りからそれぞれのテーマの地図化適性について考察を行う。ジニ係数とは、所得や資産の分布の不平等度を表す指標の 1 つである。係数は 0 と 1 の間の値で示され、完全に平等なとき最小値 0 をとり、不平等度が大きいほど 1 に近づく。イタリアの統計学者ジニ (C. Gini1884～1965) が考案した (大辞林)。

③有効写真の分布

テーマごとに有効写真のみをマッピングして、テーマ別観光マップを作成し、その分布の状況などから、実用化に向けての課題点などを考察する。

(2) 使用ツール

倉田により Web 上で一般公開されている、Flicker データ抽出ツール、可視化ツールを用いた。(http://www.comp.tmu.ac.jp/kurata/tpm/)

データの抽出ツールは、対象領域・期間・キーワード(省略可)を指定する

と、Flickr のサーバから投稿写真データを自動抽出できるツールである。可視化ツールは、背景地図の作成には Google Maps API を使用しており、地図のスケールや表示範囲の変更と、衛星写真への切り替えが可能なツールである。ヒートマップの作成には、同 API に実装されている Heatmap Layer を用いている。また、ヒートマップとしての表示だけでなく、サンプル写真を地図上に表示することもできる。

(3) 対象地域と対象期間

対象地域は東京駅を中心に東西 10km、南北 10km の範囲とし、2013 年に撮影され Flickr に投稿された写真を用いる。多様なテーマのマップの作成を試みるには、写真の投稿枚数の多い、都市部が適していると考えたので、まずは首都である東京を対象地とした。

(4) 地図化するテーマの選定

「既存の手法では地図化しにくかった観光資源、観光情報」であることを条件に、大きくわけて以下の 2 つの方法でのデータの抽出の仕方を提案する。

a.観光対象をキーワード検索によって抽出できるもの

この例として、以下の 7 項目 23 個のテーマを選定した。それぞれの選定理由を以下に述べる。

a1.夜

①夜

観光ポテンシャルマップの表示ツールには、写真の撮影時刻別に表示する機能があるが、夜間の観光ポテンシャルを見出すことが狙いであると考えられる。本研究ではまず、キーワード検索によって夜の写真を抽出しマッピングすることでこのテーマを地図化することについて考察する。

a2.気象現象

①雨、②雪、③夕焼け、④朝焼け、⑤虹

青野・倉田（2014）は、日本の観光を考えるうえで雨天への配慮は避けられない問題だとし、さらに、「雨の風情を楽しみたい」「雨ならではの自然を体験したい」など、雨の日の観光に対する肯定的な意見も見られるとの研究結果を示している。そこで、雨の日の観光のポテンシャルを定量的に分析するために、今回、雨をテーマとして選定した。また、それに加え、雪、夕焼け、朝焼け、虹、という気象現象をテーマとして選定した。

a3.野生生物

①猫、②蝶、③ホタル

屋外で見られる、観光資源となり得る野生生物である。例えば猫は、宮城県の田代島の観光客のうち、約 87%が「猫」を目的に来ている（白柳 2011）、というような事例もあることから、観光資源として注目されていることがわかる。いずれも、動き回るので通常の方法では地図化が難しいので、テーマとして選定した。

a4.気象庁が季節観測を行っている植物

①桜、②梅、③紫陽花、④銀杏、⑤楓

気象庁で行われている生物季節観測の情報の一部として、特定の植物の開花日や紅葉した日が記録されている。例えば、桜の注目度は高いため撮影する人が一定数いることが予測される。また、開花時期が重要になるので、写真に付加された撮影時刻のデータにより、時系列別のマップの作成が期待できる。同じく、紅葉（こうよう）も人気の観光資源であることから、気象庁の Web サイトで紹介されている 5 種類の植物をテーマとして選んだ。

a5.Deep Japan 2013 訪日観光キーワード 1~4 位

①そば、②寿司、③ラーメン、④着物

インバウンドの促進は、日本の観光政策における重要な施策の 1 つであるが、2014 年 1 月に、訪日外国人観光客を対象とした国内旅行 How To サイト Deep Japan³における、2013 年訪日観光キーワード BEST20 が発表された（マイナビニュース 2014 年 1 月 15 日付）。今回発表されているのは「キーワード」

であり、特定の場所などをさしているわけではないので、地図化を試みたく、ベスト 20 までは発表されているうちの 1~4 位をテーマとして選定した。5 位は「しょうゆ」で、地図化できるような観光資源ではないと考えたので外した。

a6. イベント性が高いもの

① コスプレ、② 神輿、③ クリスマスツリー

出現する時期や場所はある程度予測できるが、いつもそこにあるわけではないものとして、イベント性の高い 3 つのテーマを選定した。

a7. 広範囲から確認できる観光資源

① 富士山、② 東京スカイツリー

これらは観光資源としてメジャーで、そのものの位置は固定されているが、大きな観光資源なので、撮影地が様々になる。写真に付加された位置情報を用いて、その観光資源を見ることができる場所を地図化したく、テーマとして選定した。

b. 特定の条件下に現れるもの

① 夜間、② 雨天時

キーワード検索で抽出できるもののうち、夜と雨に関しては、撮影時刻そのものや、撮影時刻と過去の気象データとの組み合わせによって抽出が可能と考えた。そこで、これらを特定の条件下に現れるものとして、テーマとして選定した。

(5) 抽出方法と判別方法

キーワード検索による抽出は、それぞれのテーマについて、日本語と英語の両方で検索を行い、いずれかを写真のタイトルまたは写真に付加されたタグに含むものを抽出する。検索ワードはそれぞれ表 1 に示した通りである。

条件による抽出については、枚数がきわめて多くなることが予想されるため、2013 年の 5 月と 9 月に撮影された写真の中から、① 撮影時刻 20-24 時のもの、② 雨天時⁴に撮影されたものを抽出した。

判別は目視により、有効写真、不要写真、の 2 種類に判別する。テーマによっては、有効写真となりうる写真、を設け、3 種類に判別する。このようなテーマでは、テーマ名のあとに「(狭い定義)」とあるものは、有効となりうる写真を有効写真に含めず、「(広い定義)」とあるものは、有効となりうる写真を有効写真に含めている。いずれも、有効写真に当てはまらないものを「不要写真」とする。判別基準は、それぞれ表 1、表 2 に示したことが、その写真を見て客観的に確認できるかどうかである。例外的に条件による抽出の「夜間」のみ、明らかな不要写真、という項目を設けている。

a. キーワードによる抽出

表 1 各テーマの検索ワードと手動判別基準（キーワードによる抽出）

テーマ (検索ワード)	有効写真	有効写真となりうる写真
夜,night	夜空	夜空ともとらえられる空、暗さや照明などにより夜と予測できる屋外
雨,rain	空中の雨の線、傘をさしている人、水面に波紋	窓等についた雨粒、水たまり、濡れた路面
雪,snow	空中の雪の粒、(周りに雪が積もっていて) 傘をさしている人	積もった雪
夕焼け,sunset	空またはビルや水面等への太陽光の反射が写っているもののうち、空またはビルや水面等への反射の、赤み・黄色み	—
朝焼け,sunrise	空またはビルや水面等への太陽光の反射が写っているもののうち、空またはビルや水面等への反射の、赤み・黄色み	—
虹,rainbow	空にかかった虹	—
猫,cat	屋外にいる猫	—
蝶,butterfly	生きている蝶	—
ホタル,firefly	生きているホタル	—
桜,cherry blossom	開花している・開花しそうな桜の花	—
梅,plum blossom	開花している・開花しそうな梅の花	—
紫陽花, hydrangea	開花している・開花しそうな紫陽花の花	—
銀杏, ginkgo	色づいている銀杏	—
楓,maple	色づいている楓	—
そば,soba	外食の蕎麦	—
寿司,sushi	外食の寿司	—
ラーメン,ramen	外食のラーメン	—
着物,kimono	屋外にいる着物(浴衣を含む)を着た人	—
コスプレ,cosplay	コスプレをした人(屋内外は問わない)	—
神輿, portable shrine	神輿の本体、または一部	—
クリスマスツリー, Christmas tree	公共空間にある、円錐状のクリスマスツリー	—
富士山,Mt.Fuji	富士山	—
東京スカイツリー, Tokyo Skytree	スカイツリーの外観(展望台部分と、その上下の部分)	—

b. 条件による抽出

表 2 各テーマの手動判別基準（条件による抽出）

テーマ	有効写真	有効写真となりうる写真	明らかな不要写真
夜間	夜空	暗さや照明などにより夜と予測できる屋外	明らかに 20 時から 24 時ではない空
雨天時	空中の雨の線、傘をさしている人、水面に波紋	窓等についた雨粒、水たまり、濡れた路面	—

4.2 結果と考察

①データ抽出数

a. キーワード検索による抽出

抽出された写真を判別した結果、表 3 に示す通りとなった。

表 3 各テーマの手動判別結果（キーワード検索による抽出）

	有効写真	不要写真	有効写真率	合計
夜（狭い定義）	3873	2437	60.8%	6310
夜（広い定義）	5075	1235	80.4%	
雨（狭い定義）	199	453	30.5%	652
雨（広い定義）	306	346	46.9%	
雪（狭い定義）	214	584	26.8%	798
雪（広い定義）	409	389	51.3%	
夕焼け	811	79	91.1%	890
朝焼け	43	45	48.9%	88
虹	10	542	1.8%	552
猫	1359	656	67.4%	2015
蝶	33	23	58.9%	56
ホタル	0	126	0%	126
桜	3154	513	86.0%	3667
梅	191	82	70.0%	273
紫陽花	130	2	98.5%	132
銀杏	174	54	76.3%	228
楓	87	37	70.2%	124
そば	239	338	41.4%	577
寿司	413	313	56.9%	726
ラーメン	581	287	66.9%	868
着物	72	43	62.6%	115
コスプレ	1338	112	92.3%	1450
神輿	224	119	65.3%	343
クリスマスツリー	110	109	50.2%	219
富士山	51	100	33.8%	151
東京スカイツリー	1408	1475	48.8%	2883

b. 条件による抽出

抽出された写真を判別した結果、表 4 の通りとなった。

表 4 各テーマの手動判別結果（条件による抽出）

	有効写真	不要写真	有効写真率	合計
夜間（狭い定義）	1217	4267	22.7%	5484
夜間（広い定義）	1646	3838	30.0%	
雨天時（狭い定義）	384	3709	9.4%	4093
雨天時（広い定義）	679	3414	19.9%	

有効写真枚数に関する考察

a. キーワード検索による抽出において、有効写真枚数が 1000 枚以上のものを、有効写真枚数が多いテーマ、50 枚以下のものを、有効写真枚数が少ないテーマとして考察を行った。

有効写真枚数が多いテーマについて

有効写真枚数が多い順に、夜（広い定義）、夜（狭い定義）、桜、東京スカイツリー、猫、コスプレ、となった。夜は、タグとして登場することが多い。桜は、時期は限定されるものの、期間内であれば都内にスポットも多く、被写体が動かないので撮影はしやすい。東京スカイツリーは、被写体は 1 つしかないものの、大きいので、撮影地が広範囲に渡るのではと考えられる。猫、コスプレの写真を見てみると、同じ猫、同じ人と思われる写真が連続して何枚も投稿されている。同じものでも様々な表情、角度で楽しみたいタイプの観光資源なのではないかと考えられる。

有効写真枚数が少ないテーマについて

少ない順に、ホテル、虹、蝶、朝焼けとなった。ホテルは有効写真が 1 枚も抽出できなかった。都内でも、ホテル観賞のイベントが行われるなど観光資源としての注目は低くないはずだが、生息数が少ないことと、被写体が小さく暗

闇の中を飛び回ることから、撮影が難しいことが原因だと考えられる。蝶は、ホタルよりは注目度が低いと考えられるが、出会える確率も比較的高いことと、花と一緒に撮られることが多く、絵になるので被写体としてホタルよりは人気があるのでは、と考えられる。虹は発生のタイミングをはかれないので、撮影が難しいと思われる。朝焼けは、見た目は夕焼けと似ているが、通常の人々の活動時間帯ではないため、有効写真枚数は夕焼けに比べてかなり少ない。

有効写真率に関する考察

a.キーワード検索による抽出において、有効写真率 80%以上のものを、有効写真率が高いテーマ、40%以下のものを、有効写真率が低いテーマとして考察を行った。また、b.条件による抽出を行ったテーマについて

有効写真率が高いテーマについて

高いものから順に、紫陽花、コスプレ、夕焼け、桜、夜（広い定義）、となった。紫陽花、コスプレは、他のものを表すタイトルやタグとして登場することが少ない。一方、「梅」や「銀杏（いちろう）」の場合、不要写真の中には梅料理や銀杏（ぎんなん）料理の写真が含まれるなど、キーワードだけでは目当てのものを絞り込めないで、有効写真率が低下する。夕焼けは、朝焼けに比べて有効写真率が高い。この差の理由は、朝焼けの写真には「サンライズ」というお店で撮られた不要写真が多く含まれていたためである。夜（広い定義）は、まず条件による抽出と比べると、わざわざタイトルやタグにキーワードを含めているだけあって、「有効となりうる写真」まで含めれば有効写真率が 80%を超える。

有効写真率が低いテーマについて

低い方から順に、ホタル、虹、雪（狭い定義）、雨（狭い定義）、富士山、となった。ホタルは、先にも述べたように、有効写真を 1 枚も抽出することができなかった。そして、不要写真のほとんどは、「ホタル」と名のついたイルミネーションのイベントの写真であった。虹の場合は、「レインボーブリッジ」の写真が不要写真として多く出現した。富士山は、全く関係ないと思われる写真に、

他の有名な観光地名等と併せてタグが付けられているパターンが見られた。この原因としては、一連の旅行写真全てに、その旅行で訪れた地名をタグとして付けて投稿した可能性が考えられる。東京スカイツリーにも、ややこの傾向が見られる。雪（狭い定義）、雨（狭い定義）は、現在進行形で降雪・降水があることが有効写真の条件となるが、それを写真の目視で判別するのは難しく、実際には降っていても不要と判断した写真もあるかもしれない。

b.条件による抽出について

b.条件による抽出は、いずれのテーマも有効写真率が低い。有効写真を目視で客観的に判別しているため、単に条件で抽出したのでは、こちらの意図するものがうまく抽出できない。例えば、夜間に撮影された飲食店での写真は、実際には夜間の行動を表し、夜間観光のポテンシャルを示す写真なのかもしれないが、屋内の写真は目視による時間帯の判別ができないため、本研究においては不要写真となってしまふ。また、夜間に撮られたはずの写真のうち、9.6%は明らかに撮影時刻がずれていると判別できる写真（明らかな不要写真）であったことから、写真に付加されている撮影時刻のデータを完全に信頼することもできない。雨天時の場合も夜間同様、屋内の写真が多いことに加え、先の考察にもあったように雨が降っていることを客観的に判断するのは難しく、有効写真率は夜間よりもさらに下がる。また、撮影時刻の信頼性の問題に加え、気象データの信頼性の問題もあり、有効な写真を抽出するのがより難しいと考えられる。いずれのテーマも、キーワード検索の場合は、1年間あたりの有効写真枚数は減少するが、有効写真率が向上する。有効写真のみをうまく選別するルールを構築することができなければ、キーワード検索による抽出の方が、純度の高いデータを得ることができるということである。

②投稿枚数の偏り

各テーマの、有効写真の投稿者とそれぞれの投稿枚数のデータから、ジニ係数を求めた。結果は表 5 に示した通りである。

表 5 各テーマのジニ係数

テーマ	ジニ係数	投稿者数	有効写真枚数
夜（狭い定義）	0.66	837	3873
夜（広い定義）	0.70	926	5075
雨（狭い定義）	0.41	99	199
雨（広い定義）	0.43	141	306
雪（狭い定義）	0.54	65	214
雪（広い定義）	0.61	109	409
夕焼け	0.60	256	811
朝焼け	0.28	29	43
虹	0.09	9	10
猫	0.90	96	1359
蝶	0.36	15	33
ホテル	-	0	0
桜	0.71	431	3154
梅	0.53	54	191
紫陽花	0.53	40	130
銀杏	0.68	40	174
楓	0.49	36	87
そば	0.72	51	239
寿司	0.65	105	413
ラーメン	0.54	169	581
着物	0.35	43	72
コスプレ	0.90	28	1338
神輿	0.69	25	224
クリスマスツリー	0.61	27	110
富士山	0.28	33	51
東京スカイツリー	0.66	330	1408
夜間（狭い定義）	0.65	266	1217
夜間（広い定義）	0.67	313	1646
雨天時（狭い定義）	0.64	83	384
雨天時（広い定義）	0.67	135	679

ジニ係数に関する考察

ジニ係数が 0.4 以下のものを、投稿枚数に偏りが少ないテーマ、0.7 以上のものを、投稿枚数の偏りが大きいテーマとして考察を行った。

投稿枚数に偏りが少ないテーマについて

ジニ係数の低い順に、虹、朝焼け、富士山、着物、蝶となった。ホテルを除く、有効写真枚数ワースト 1～5 位に含まれるテーマと一致する。総じて撮影の機会も少なく、ベストショットを捉えにくいものであると考えられる。虹を狙って撮影することは難しく、着物も、道を通行している人の後ろ姿をとらえたような写真も多いのであまり枚数も多くならない。富士山は、東京から見ると場所も限られ、天候の条件がそろふことも必要で、そこまで大量に撮影する人がいないのだと考えられる。

投稿枚数の偏りが大きいテーマについて

ジニ係数の高い順に、猫、コスプレ、そば、桜となった。特に高い猫とコスプレというのは、一部の人がそのテーマの写真を多く投稿しているということであり、熱狂的なファンがいるであると考えられる。この 2 つのテーマに関しては、投稿している人は大量にいるが、全ての人が複数枚投稿しているわけではなく、一部の熱狂的なファンによって有効写真枚数を伸ばしているテーマだと言える。そばは、ラーメンや寿司に比べてジニ係数が高い。熱狂的なファンという、ラーメンの方が高いような気がするが、写真を投稿する料理としてラーメンはもはや一般的になっていると考えられる。桜は、意外にもジニ係数が高く、猫やコスプレ同様、様々な角度から楽しみたいという思いで写真を撮り、複数枚投稿している人々と、一般的なものとして何気なくアップする人々の両者が存在しているのではと考えられる。

③有効写真の空間分布

それぞれのテーマにおいて、有効写真のみをマッピングした理想の観光マップを図 2 から図 35 に示した。表示の際、カーネル半径（各点の周囲に密度分布が生じる範囲）は 500m、可視化上限（それ以上の値だと一律同色となる閾値）はマッピングする有効写真の平方根に設定した。また、本研究は、「テーマ別観光マップ」の研究であり「観光ポテンシャルマップ」の研究ではないが、考察の際には、各テーマの有効写真の分布に対し「ポテンシャル」という言葉を用いることにする。ポテンシャルとはすなわち、撮影密度分布のことである。

夜（キーワードベース）、夜間（時間ベース）

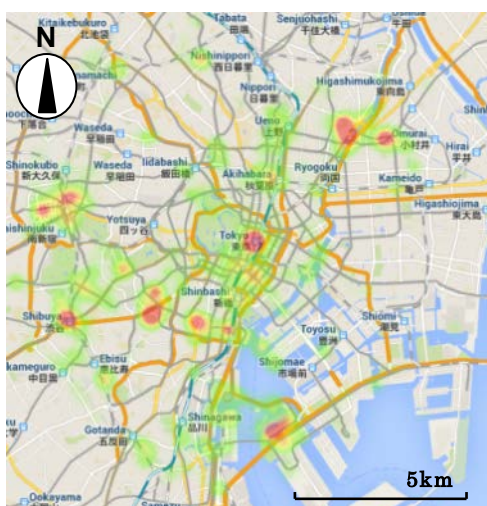


図 2 夜（狭い定義）のマップ



図 3 夜（広い定義）のマップ

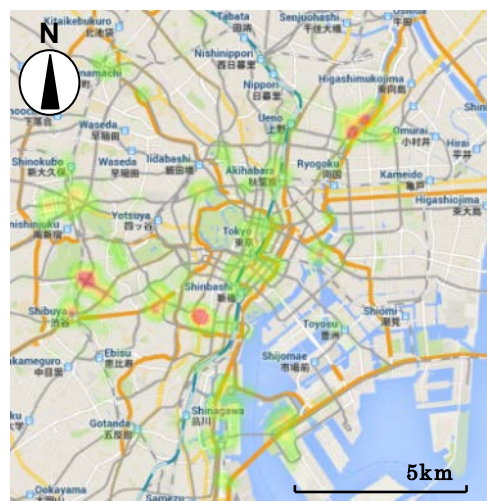


図 4 夜間（狭い定義）のマップ

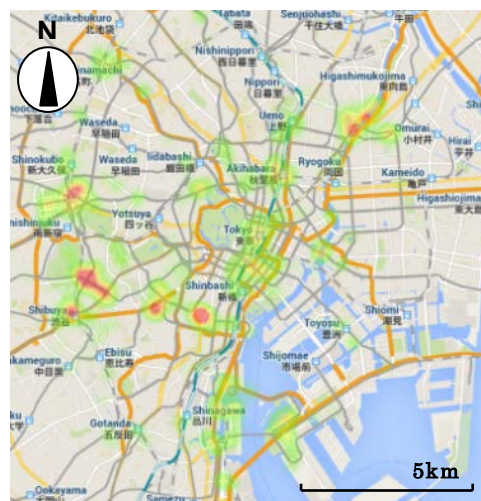


図 5 夜間（広い定義）のマップ

どちらの抽出方法でも、狭い定義と広い定義では大きな違いは見られないので、広い定義のマップ（図 3 と図 5）で考察を進める。

夜（キーワードベース）のポテンシャルが特に高いのは、東京駅、浅草、お台場、六本木、渋谷、新宿である（図 3）。夜間（時間ベース）は、六本木、原宿、浅草にポテンシャルが現れている（図 5）。大きな違いは、夜の方では大きなポテンシャルが現れている東京駅周辺が、夜間では目立たなくなっていることである。東京駅周辺の写真としては、ライトアップされた東京駅の駅舎の写真が多く見られる。夜間では、確実に夜の写真を得るために 20 時から 24 時の写真を抽出したが、この東京駅駅舎のライトアップが行われているのが、日没から 21 時までの間であったために、漏れてしまったのだと考えられる。このことから、日没時刻は季節によって大きく変わるため、条件による抽出を行う場合は、単に時間を指定するだけでなく、その時期の日没の時刻を考慮することも大切であると示唆される。また、あまり遅い時間帯であると、外出している人が少なく、不要写真の割合が多くなるという問題もある。

雨（キーワードベース）、雨天時（時間ベース）



図 6 雨（狭い定義）のマップ

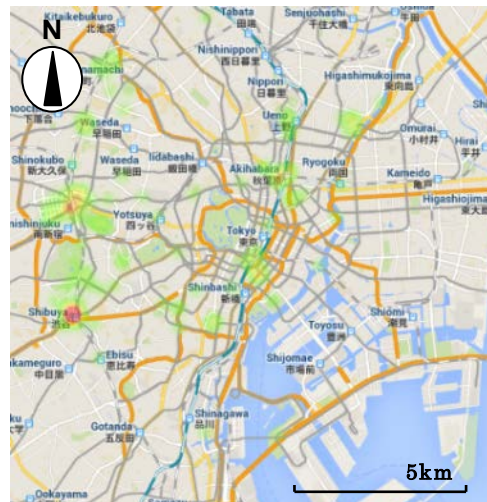


図 7 雨（広い定義）のマップ

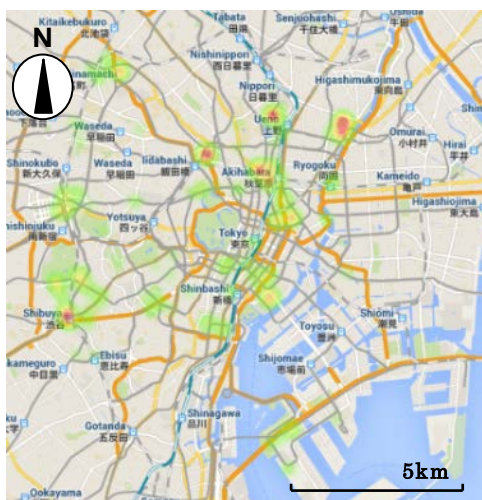


図 8 雨天時（狭い定義）のマップ

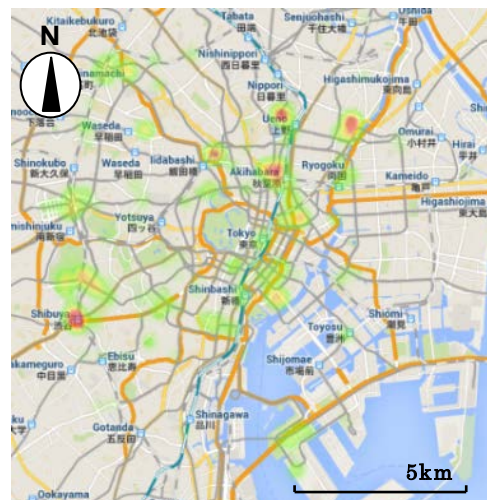


図 9 雨天時（広い定義）のマップ

どちらの抽出方法でも、狭い定義と広い定義では大きな違いは見られないので、広い定義のマップ（図 7 と図 9）で考察を進める。

雨（キーワード検索による抽出）のポテンシャルは、渋谷、新宿に見られる（図 7）。雨天時（条件による抽出）は、渋谷、浅草、上野、後樂園、秋葉原、にポテンシャルが見られる（図 9）。共に目立つのは渋谷で、小石川後樂園は、雨天時のマップでのみポテンシャルとして表れている。

これだけでは、特徴は見いだせず、雨のポテンシャルを計るには、抽出された写真を見たり、コメントを分析したりと、さらなる分析や考察が必要である。

雪（狭い定義）、雪（広い定義）、虹、朝焼け、夕焼け



図 10 雪（狭い定義）のマップ

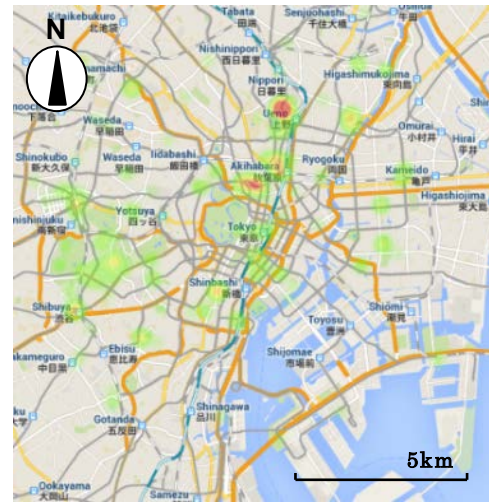


図 11 雪（広い定義）のマップ

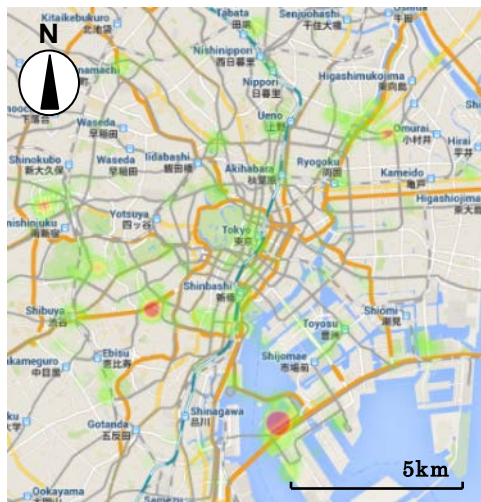


図 12 夕焼けのマップ

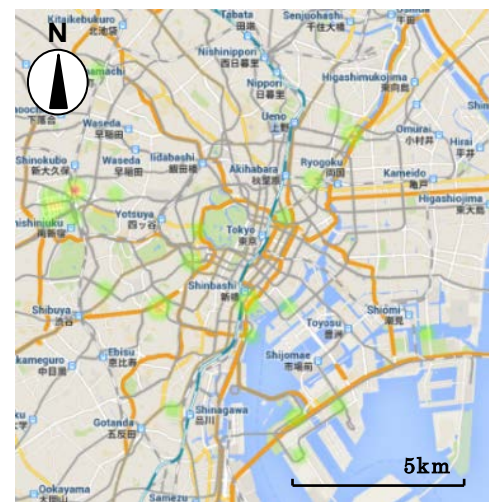


図 13 朝焼けのマップ

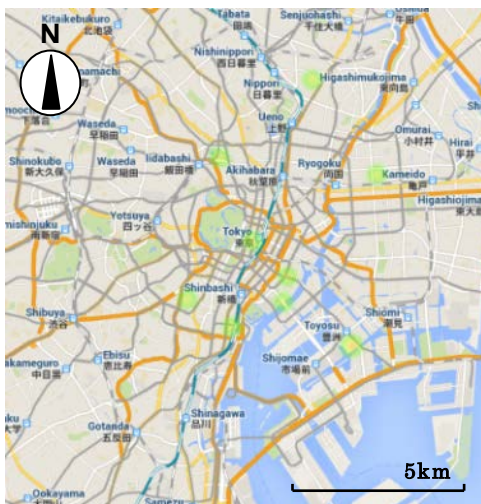


図 14 虹のマップ

雪も、狭い定義と広い定義では大きな差は見られない。目立つポテンシャルは上野に見られる（図 10、図 11）。特に、上野動物園内で撮られた写真が多い。上野動物園には、元々雪国に生息している動物も多く、雪の日でこそ見られる姿もあるのだという（東京ゾーンネット Web サイトより）。

夕焼けは、お台場と六本木にポテンシャルが見られるが（図 12）、朝焼け、虹は、枚数が少ないこともあるが、特徴が見出せない（図 13、図 14）。

猫，蝶

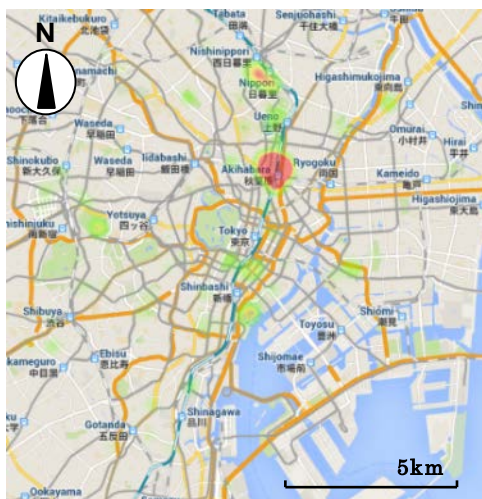


図 15 猫のマップ

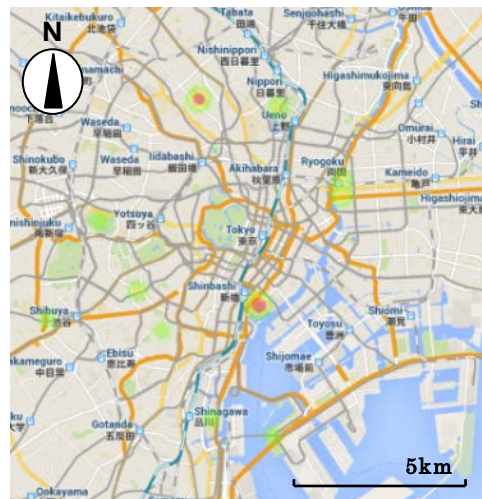


図 16 蝶のマップ

猫は、秋葉原周辺に大きなポテンシャルがある。次いで高いポテンシャルを示すのが、谷根千と呼ばれる、猫が多くいることで有名な地域である（図 15）。また、秋葉原駅周辺を拡大し、表示の設定を変える（半径 100m）と、図 17 のようなマップが得られる。このように表示することで、特にどのようにポテンシャルが分布しているかを見る事が出来る。

蝶も、大きくは 2 つ、浜離宮と東京大学附属植物園にポテンシャルがある（図 16）。



図 17 猫のマップ（秋葉原駅周辺）

桜, 梅, 紫陽花, 銀杏, 楓



図 18 桜のマップ



図 19 梅のマップ



図 20 紫陽花のマップ



図 21 銀杏のマップ



図 22 楓のマップ

桜は、新宿御苑、中目黒駅周辺の目黒川沿い、皇居隣接の北の丸公園、上野公園、代々木公園に目立つポテンシャルがある。また、中目黒駅から北西に带状にポテンシャルが見られる(図 18)。これは、目黒川沿いの桜だと思われる。梅のポテンシャルは、湯島天満宮と、錦糸公園に。紫陽花は、浜離宮、清澄庭園、新宿御苑(図 19)。銀杏は、明治神宮外苑(図 20)。楓は、小石川後樂園、六義園であり(図 21)、いずれも、これらの植物を観光資源に、注目を集めている場所である。このマップだけでは一般的な結果が出たにすぎない。

そこで桜を例に、時期をわけてマップを作成した。3 月、4 月、それ以外の月、の 3 種類のマップを次ページの図 23 から図 25 に示す。(比較のために、年間の桜のマップ(図 18)も再び掲載している。)それを見ると、3 月のマップ(図 23)は年間のマップ(図 18)とほぼ同じである。他の 2 枚(図 24、図 25)を見ると、新宿御苑は変わらずポテンシャルを保つのに比べ、他の場所、特に目黒川沿いのポテンシャルが低くなる。また、一般的に東京における桜の見頃は 3、4 月だが、図 25 を見るとわかるように、それ以外の時期に咲く桜もある。さらに時期を細かく区切ってマップ表示することで、桜の開花状況を地図化できると考えられる。現状では、月ごとの表示が限界である。他に注目すべき点としては、図 25 において新宿御苑の次にポテンシャルがある亀戸中央公園だが、2013 年の 10 月 9 日に日本全国で記録的な暑さを記録しており、亀戸中央公園のソメイヨシノが季節外れの開花をしたということがあった(日テレ NEWS24 2013 年 10 月 9 日付け)。サンプルデータの抽出期間が限定的だと、このような例外が目立ちやすくなる。

桜以外のテーマに関しても、桜の例のように、時期別のマップを作ることで、有用なものになると予想できる。

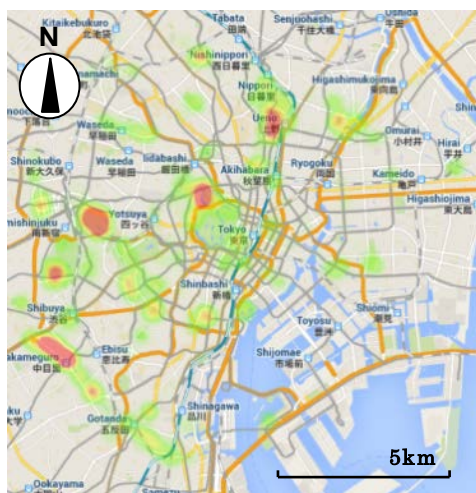


図 18 桜のマップ (年間)

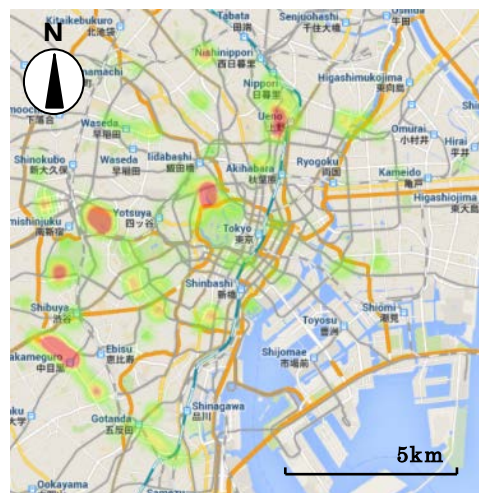


図 23 桜 3 月のマップ



図 24 桜 4 月のマップ



図 25 桜 3,4 月以外のマップ

そば、寿司、ラーメン、着物

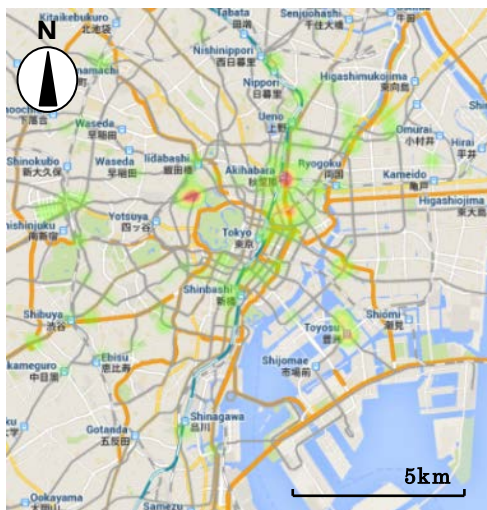


図 26 そばのマップ

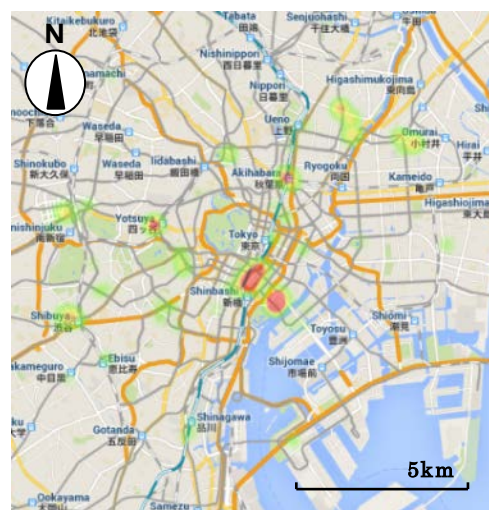


図 27 寿司のマップ

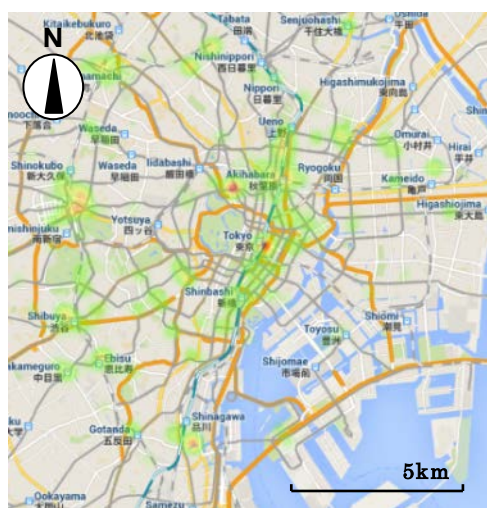


図 28 ラーメンのマップ

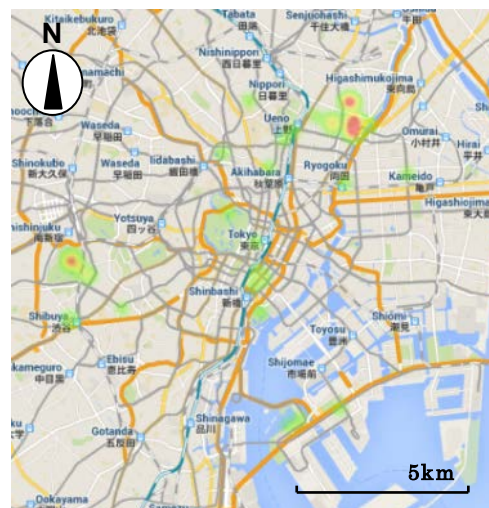


図 29 着物のマップ

着物は、浅草周辺と明治神宮にポテンシャルがある（図 29）。そばは、靖国神社周辺と、秋葉原駅周辺にポテンシャルがある（図 26）。寿司は、築地周辺と銀座周辺に大きなポテンシャルがある（図 27）。ラーメンは、東京駅や水道橋駅付近にやや高めのポテンシャルが見られるものの、分散している。それだけラーメン店が都内に多いということはわかるが、これだけではマップとしての有用性には欠ける。そばや寿司に比べて、ラーメンの味付けには幅があるということもあり、「ラーメン」という一括りのテーマではなく、さらに細分化が求められる。現時点での対応としては、マップの表示ツールの仕様として写真自体を地図上に表示できるので（図 30）、その写真を参考にするのである。

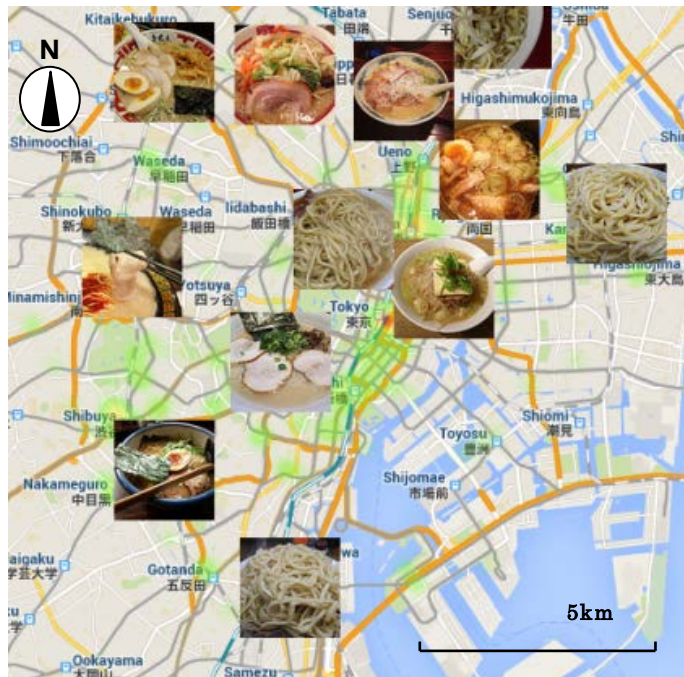


図 30 ラーメンのマップ（写真表示版）

コスプレ、神輿、クリスマスツリー

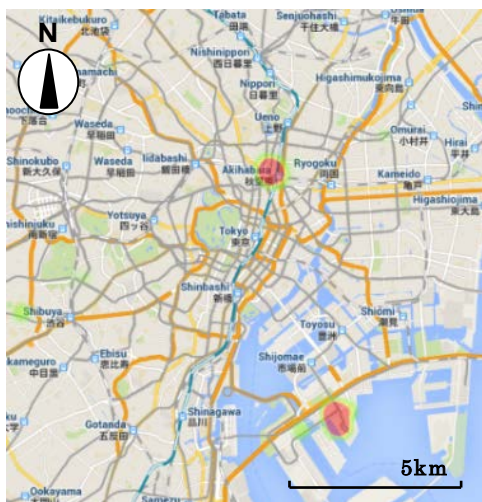


図 31 コスプレのマップ

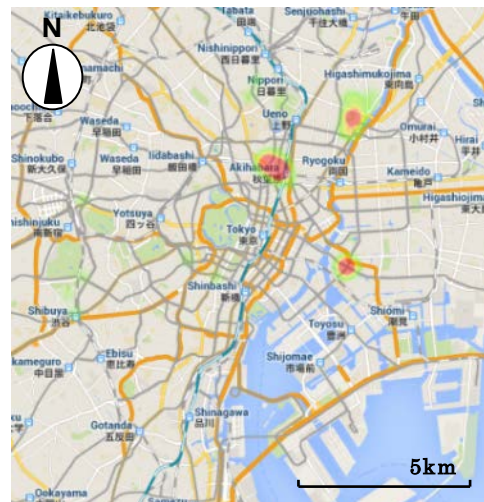


図 32 神輿のマップ

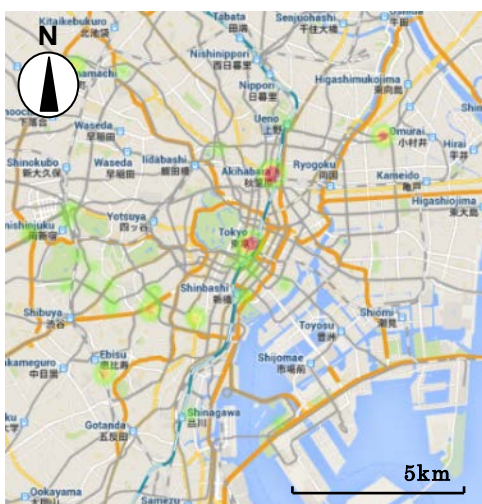


図 33 クリスマスツリーのマップ

コスプレは、秋葉原と東京ビッグサイトに大きなポテンシャルがある（図 31）。神輿は、秋葉原、門前仲町、浅草の三地区、クリスマスツリーは、秋葉原、東京駅、東京スカイツリー周辺の三地区でポテンシャルが高い（図 32、図 33）。これらは、イベント性の高いテーマなため、それぞれ時期が限定されるので、地図として提供する際には、見ることのできる時期の表示などが必要である。

富士山，東京スカイツリー

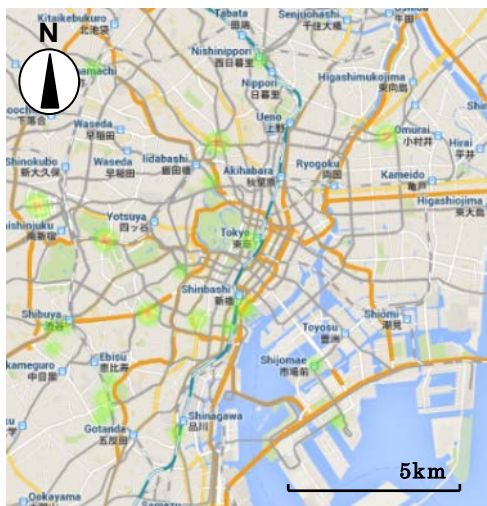


図 34 富士山のマップ



図 35 東京スカイツリーのマップ

富士山は、ポテンシャルが点在しており、都内には目立った鑑賞スポットはないように思える。しかしながら、ポテンシャルがやや高めに出ているところは、文京シビックセンター、東京スカイツリー、東京都庁、六本木ヒルズなどの展望台のある施設の周辺であることがわかるので、このマップは主に展望スポットのポテンシャルが表されている（図 34）。東京スカイツリーは、対象物周辺と、墨田川沿いにポテンシャルが高く出ている（図 35）。それ以外の場所だと、富士山のポテンシャルと一致するところが多いので、展望施設からの撮影だと思われる。これらのテーマに関しても、写真を表示し、それを見ながら記念撮影の場所を選ぶというような使い方ができそうだ。

4.3 調査 1 のまとめ

この調査によってまず、テーマごとに有効写真を手動で判別することで、解釈のできるマップを作成することができた。しかし、テーマによって、写真抽出枚数、有効写真率、手動での判別の難易度の違いがあるなど、課題は多い。

そして、以上のことから、キーワード検索を行って写真を抽出するだけではそのテーマに関するマップを作成することは難しく、地図化したいテーマの性質を考える必要があることがわかった。そのような点を踏まえて、この方法で作成された地図が、「そのテーマに関する写真撮影活動の分布」であることを理解して扱うことが必要である。

V. 調査 2 自動選別ルール構築

5.1 調査の方法

(1) 概要

調査 2 では、先に挙げたテーマの一部に対して、有効写真を自動選別するためのルールの構築を試みる。このようなルールを構築する理由としては、今後、他地域や他期間のマップを作成するために、より大量のデータを処理する必要性が出てきたときに、人力では判別が追いつかなくなると予測されるからである。ルールの構築には、決定木学習法を用いる。得られたルールは、東京におけるそれぞれのテーマの全写真のデータと、大阪における同一テーマの写真のデータに適用し、ルールの有効性を検証する。ただし、東京スカイツリーは東京のデータへの適用に限る。具体的な分析の内容は、以下の通りである。

①ルールの構築

テーマごとに、ルール構築に用いる説明変数を選出し、ルールを構築したのち、そのルールの解釈を行う。

②ルールの適用

適用前、適用後のデータ数などからルールの有効性を考察する。

(2) ルール構築を行うテーマ

a. キーワード検索による抽出で有効写真が 500 枚以上のテーマについて、自動選別ルールの構築を行う。その条件に当てはまるのは、夜（狭い定義、広い定義）、夕焼け、猫、桜、ラーメン、東京スカイツリー、コスプレの 7 テーマである。しかしながら、夕焼けとコスプレは不要写真がそれぞれ 79 枚、112 枚と、決定木を算出するにはサンプルデータが少なく（詳しくは後述する）、すでに有効写真率が 90%を超えていることもあり、今回はルール構築を行うテーマからは外し、残りの 5 テーマについてルールの構築を行う。

(3) 決定木学習法

与えられた写真が各テーマにおける観光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。それぞれのテーマにおいて決定木を算出するための学習データには、同数の有効写真と不要写真を使用する、すなわち正負事例のバランスをとる必要があるため、少ない方の枚数に合わせる。各テーマにおける、サンプル枚数は表 4 の通りである。今回、決定木算出には IBM SPSS Decision Trees 20 を利用する。学習アルゴリズムとして Exhaustive CHAID (Biggs et al. 1991) を使用し、ツリーの最大深さは 5、各ノードの最小事例数は、夜（狭い定義・広い定義）と東京スカイツリーは 100、データ数の少ない猫と桜とラーメンは 50 に設定する。

表 6 決定木学習法に用いるサンプルデータ数

テーマ	有効写真	不要写真
夜（狭い定義）	全 3873 枚から無作為抽出した 2437 枚	全 2437 枚
夜（広い定義）	全 5075 枚から無作為抽出した 1235 枚	全 1235 枚
猫	全 1359 枚から無作為抽出した 656 枚	全 656 枚
桜	全 3154 枚から無作為抽出した 513 枚	全 513 枚
ラーメン	全 581 枚から無作為抽出した 287 枚	全 287 枚
東京スカイツリー	全 1408 枚	1475 枚から無作為抽出した 1408 枚

(4) 説明変数

説明変数の候補としては、①タイトルの有無、②タグの有無、③ファイルソースの有無、④写真投稿者の Flickr への居住地登録の有無、⑤撮影地から居住地までの距離（50km 圏内、500km 圏内、500km 圏外、データ無）、⑥撮影機器の機器メーカー（Apple 社製、他社製、データ無）、⑦撮影時刻（24 段階）、に加え、⑧撮影月（12 段階）、⑨シャッタースピード（0-1/100 秒、1/100-1/10 秒、1/10-1 秒、1-10 秒、10-100 秒、100 秒以上、データ無）、⑩撮影地から対象物までの距離（0-10m、10-100m、100-1000m、1000-10000m、10000m 以上）を用意した。

このうち、先行研究にはなかった説明変数として今回新たに加えたのは、②タグの有無、⑧撮影月、⑨シャッタースピード、⑩撮影地から対象物までの距離である。②タグの有無は、先行研究にあった「タグ数」をアレンジし、シンプルにした。⑧撮影月は、テーマによっては季節性のあるものもあるので、選別の鍵になると考えた。⑨シャッタースピードは、特に夜の写真の選別に鍵になると考えた。⑩対象物と撮影地の距離は、対象物が特定され、固定されているテーマにのみ使える可能性のある変数である。本研究においては、東京スカイツリーの写真の選別の鍵になると考えた。

一方、タグ数、被写体距離レンジ有無、同じ撮影者が同日に撮影した写真群における相対位置、という説明変数は、先行研究では用いられていたが、本研究では外したものである。被写体距離レンジは、記録率が著しく低いため、使用しなかった。また、本研究ではキーワード検索によってデータを抽出しているため、同日写真群における相対位置は求めることができないので、候補から外した。

各テーマの決定木算出に用いる説明変数を選出するために、①タイトルの有無、②タグの有無、③ファイルソースの有無、④居住地登録の有無、は、有効写真、不要写真それぞれにおける記録率を表に示して比較する。また、⑤撮影地から居住地までの距離、⑦撮影時刻、⑧撮影月、⑨シャッタースピード、そして東京スカイツリーに対してのみ⑩撮影地から対象物までの距離、は、有効写真、不要写真それぞれにおける構成比をグラフに示して比較する。⑥機器メーカーは、Apple 社製と他社製で撮影された写真それぞれにおいて、不要写真の割合を表に示して比較する。

5.2 結果と考察

①-1 ルールの構築：夜（狭い定義）

それぞれの説明変数候補における、有効写真と不要写真の比較を、表 7 と表 8、図 36 から図 39 に示した。

表 7 属性値の記録率（夜（狭い定義））

	属性値	有効写真	不要写真
メタデータ側	タイトル	98.6%	99.4%
	タグ	85.2%	88.2%
	撮影日時	100.0%	100.0%
	居住地	73.5%	66.1%
Exifデータ	ファイルソース	38.5%	44.3%
	機器メーカー	74.2%	73.7%
	シャッター		
	スピード	73.4%	71.8%

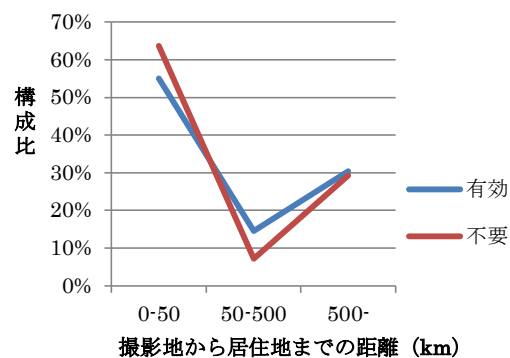


図 36 撮影地から居住地までの距離（夜（狭い定義））

表 8 機器メーカーごとの不要写真率（夜（狭い定義））

	写真枚数	不要写真枚数	不要写真率
Apple	365	123	33.7%
その他	4305	1674	38.9%
総計	4670	1797	38.5%

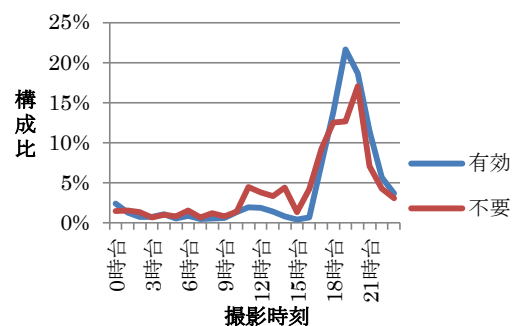


図 37 撮影時刻（夜（狭い定義））

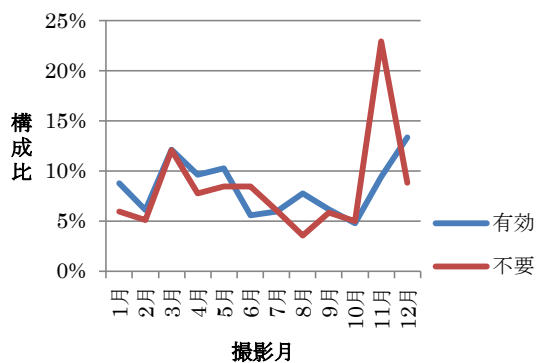


図 38 撮影月（夜（狭い定義））

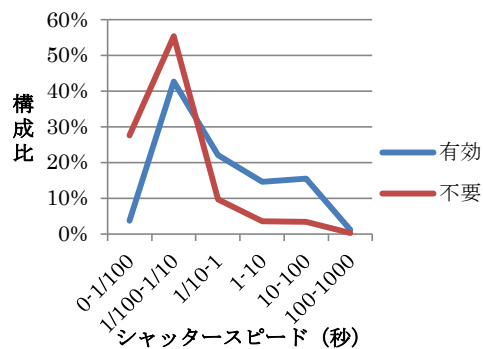


図 39 シャッタースピード（夜（狭い定義））

決定木

傾向に差が見られたのは、③ファイルソースの有無、④居住地登録の有無、⑤撮影地から居住地までの距離、⑦撮影時刻、⑨シャッタースピードの 5 つの説明変数である。この 5 つの説明変数を用いて得られた決定木を図 40 に示し、その解釈を行った。

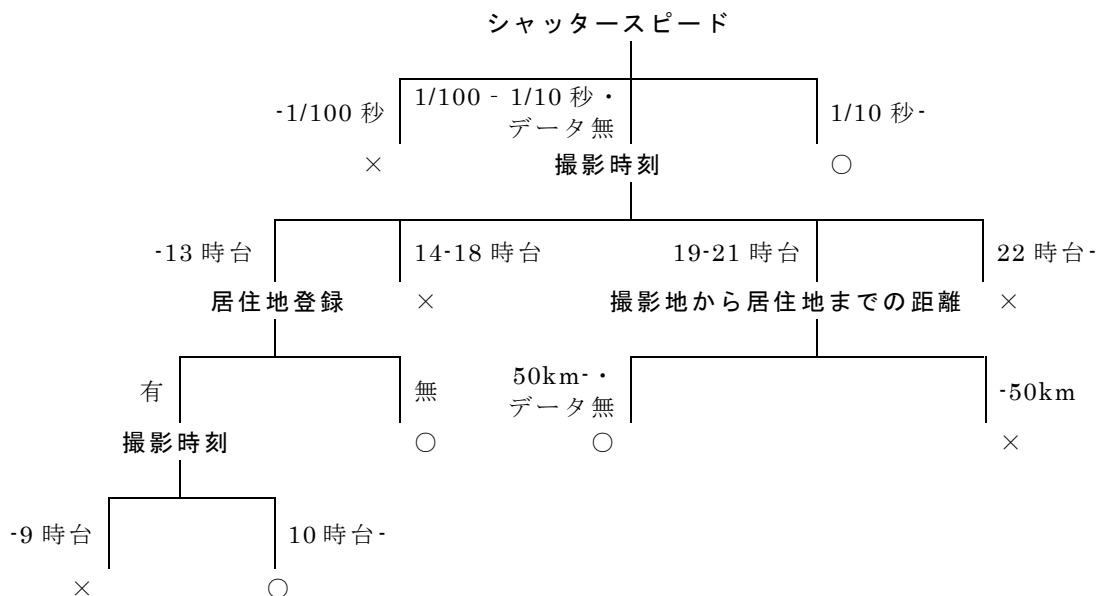


図 40 夜（狭い定義）写真選別の決定木

第一の選別の鍵は、シャッタースピードである。シャッタースピードが 0.01 秒未満の写真は、不要写真として選別される。0.1 秒以上の写真は、有効写真として選別される。夜は暗いので、撮影の際に多くの光を必要とし、シャッタースピードが長くなるためであると考えられる。0.01 秒から 0.1 秒の写真と、シャッタースピードのデータが無い写真に関しては、まず撮影時刻によって選別ができる。14 時台から 18 時台に撮影された写真、22 時台以降に撮影された写真は、不要として選別される。日中や、深夜に撮影された写真は、不要写真の可能性が高いということである。また、決定木の左側を見ると、日中に撮られたはずの写真が有効写真として選別されており、一部の写真において実際に撮影された時刻とは異なる時刻が記録されている可能性が高いと考えられる。それ以上の解釈は難しいが、撮影時刻のデータを完全に信頼することはできないということを示唆している。

①-2 ルールの構築：夜（広い定義）

それぞれの説明変数候補における、有効写真と不要写真の比較を、表 9 と表 10、図 41 から図 44 に示した。

表 9 属性値の記録率（夜（広い定義））

	属性値	有効	不要
メタデータの側面	タイトル	98.7%	99.8%
	タグ	86.4%	86.2%
	撮影日時	100.0%	100.0%
	居住地	75.9%	49.0%
Exifデータの側面	ファイルソース	39.0%	47.8%
	機器メーカー	73.6%	76.0%
	シャッター	72.2%	75.1%
	スピード	72.2%	75.1%

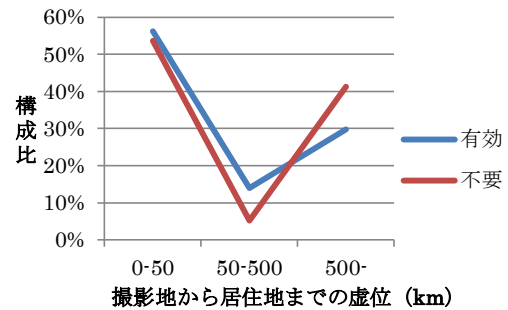


図 41 撮影地から居住地までの距離（夜（広い定義））

表 10 機器メーカーごとの不要写真率（夜（広い定義））

	写真枚数	不要写真枚数	不要写真率
Apple	365	105	28.8%
その他	4305	833	19.3%
総計	4670	938	20.1%

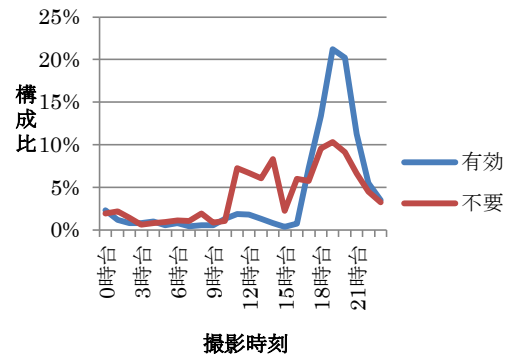


図 42 撮影時刻（夜（広い定義））

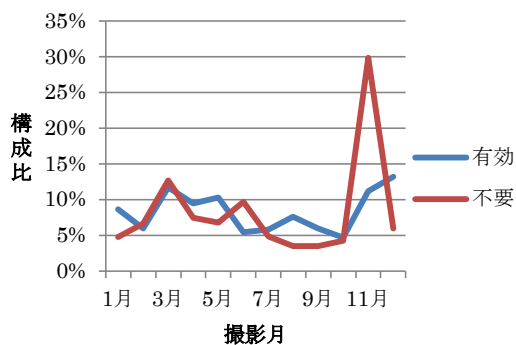


図 43 撮影月（夜（広い定義））

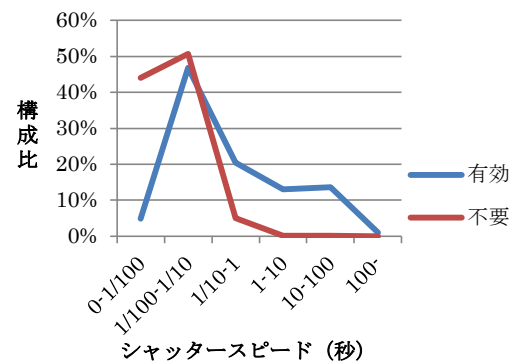


図 44 シャッタースピード（夜（広い定義））

決定木

傾向に差が見られたのは、③ファイルソース有無、④居住地登録有無、⑤撮影地から居住地までの距離、⑥機器メーカー、⑦撮影時刻、⑨シャッタースピード、の 6 つの説明変数である。この 6 つの説明変数を用いて得られた決定木を図 45 に示し、その解釈を行った。

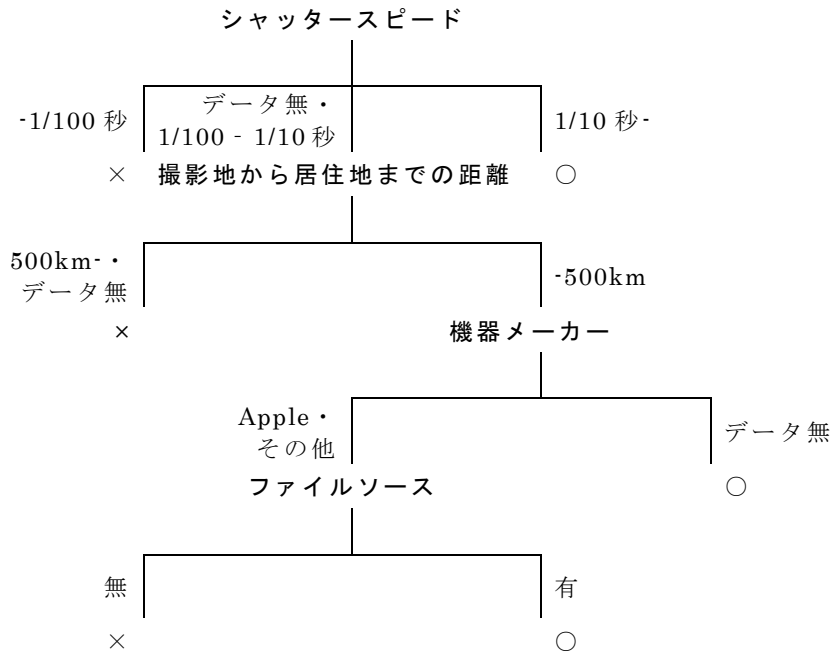


図 45 夜（広い定義）写真選別の決定木

夜（狭い定義）と同様、シャッタースピードが選別の第一の鍵である。シャッタースピードが 0.01-0.1 秒、もしくはデータの無い写真に関しては、撮影地から居住地までの距離が鍵となる。500km 圏外に住む人、居住地データの無い人の写真は不要となり、500km 圏内に住む人の写真の中で、不要と選別されているものは、スマートフォンによって撮影された写真だと考えられる。（機器メーカーが記録されていて、デジタルカメラではない）スマートフォンによって撮影されたというのが不要写真の選別の鍵になるという傾向は、従来の観光ポテンシャルマップにおける、旅行中に撮られた写真か否かの判別の場合と同様である。

①-3 ルールの構築：猫

それぞれの説明変数候補における、有効写真と不要写真の比較を、表 11、表 12、図 46 から図 49 に示した。

表 11 属性値の記録率（猫）

	属性値	有効	不要
Flicker 側の メタ デー タ	タイトル	99.7%	98.9%
	タグ	89.0%	86.4%
	撮影日時	100.0%	100.0%
	居住地	11.1%	54.6%
Exif デー タ	ファイル ソース	90.9%	41.8%
	機器メーカー	96.3%	76.1%
	シャッター スピード	96.1%	76.2%

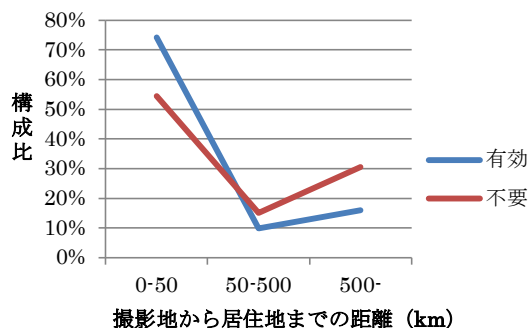


図 46 撮影地から居住地までの距離（猫）

表 12 機器メーカーごとの不要写真率（猫）

	写真枚数	不要写真枚数	不要写真率
Apple	161	142	88.2%
その他	1649	358	21.7%
総計	1810	500	27.6%

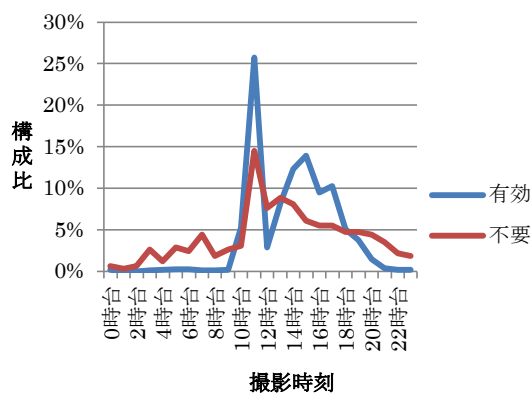


図 47 撮影時刻（猫）

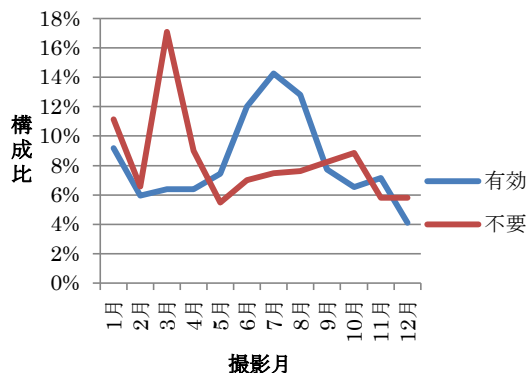


図 48 撮影月（猫）

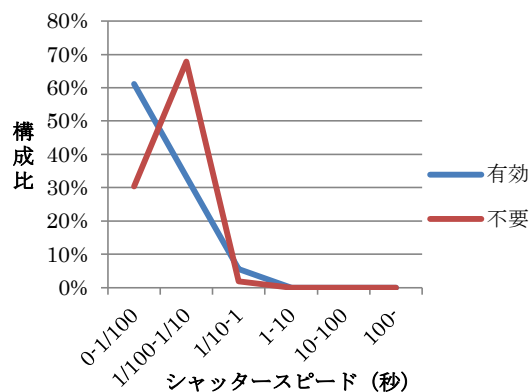


図 49 シャッタースピード（猫）

決定木

傾向に差が見られたのは、③ファイルソース有無、④居住地登録有無、⑤撮影地から居住地までの距離、⑥機器メーカー、⑦撮影時刻、⑧撮影月、⑨シャッタースピード、の7つの説明変数である。この7つの説明変数を用いて得られた決定木を図50に示し、その解釈を行った。

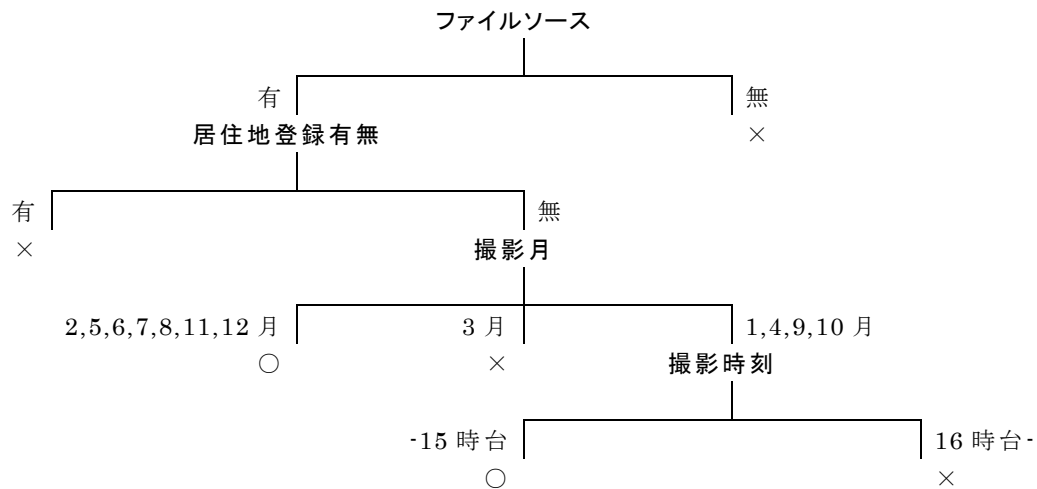


図 50 猫写真選別の決定木

選別の第一の鍵は、ファイルソースの有無である。ファイルソースが有ということは、デジタルカメラで撮影されたということである。まず、ファイルソースの無いものは、不要と選別される。猫の写真は、デジタルカメラを用いて撮影されることが多いということである。次に、ファイルソースの有るもののうち、撮影者の居住地登録データが有るものは、不要と選別される。居住地の情報、つまり個人情報を公開していない人の方が、猫を撮影する傾向にあるということだろう。さらにその後は、撮影月や撮影時刻によって選別されるが、3月に不要写真が多い理由は、写真の内容からも、見いだせなかった。撮影時刻に関しても、猫の特徴をふまえた上での解釈は難しい。

①-4 ルールの構築：桜

それぞれの説明変数候補における、有効写真と不要写真の比較を、表 13、表 14、図 51 から図 54 に示した。

表 13 属性値の記録率（桜）

属性値		有効	不要
Flickr 側の メタデータ	タイトル	96.6%	99.6%
	タグ	81.7%	93.0%
	撮影日時	100.0%	100.0%
	居住地	57.3%	56.7%
Exif 側の メタデータ	ファイル ソース	44.7%	37.2%
	機器メーカー	83.1%	74.9%
	シャッター スピード	81.6%	74.3%

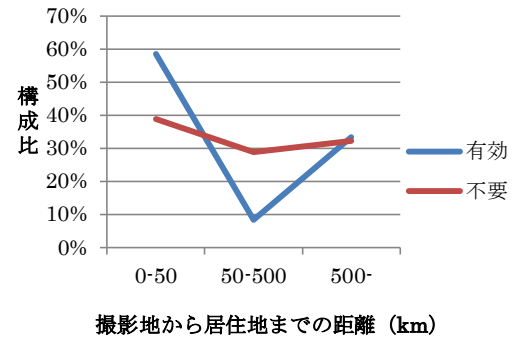


図 51 撮影地から居住地までの距離（桜）

表 14 機器メーカーごとの不要写真率（桜）

	写真枚数	不要写真枚数	不要写真率
Apple	455	42	9.2%
その他	2551	342	13.4%
総計	3006	384	12.8%

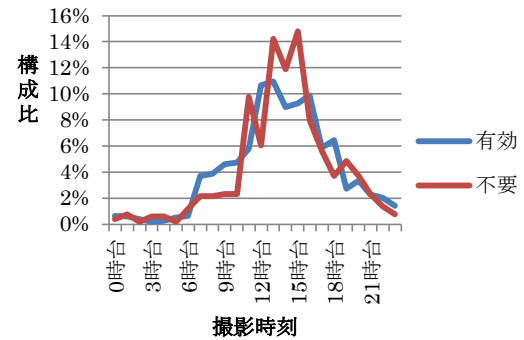


図 52 撮影時刻（桜）

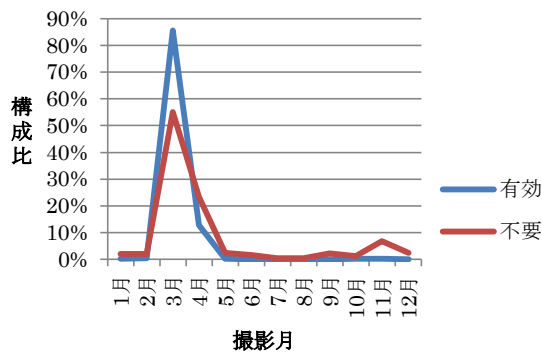


図 53 撮影月（桜）

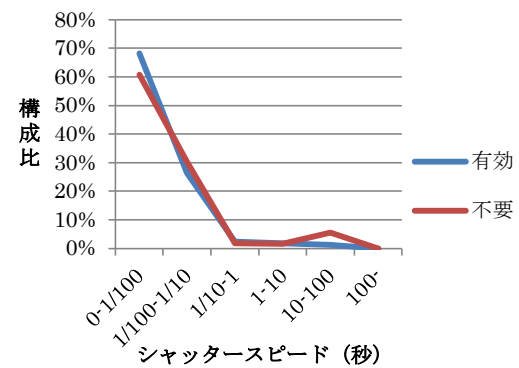


図 54 シャッタースピード（桜）

決定木

傾向に差が見られたのは、②タグの有無、③ファイルソース有無、⑤居住地と撮影地の距離、⑥機器メーカー、⑦撮影時刻、⑧撮影月の、6つの説明変数である。この6つの説明変数を用いて得られた決定木を図55に示し、その解釈を行った。



図 55 桜写真選別の決定木

第一の選別の鍵は、撮影月である。一般的に東京における桜の開花時期であると思われる3月と4月以外の時期に撮影された写真は、不要と選別された。また、3月の写真のうち、50km圏外の人による写真は不要と選別されていることから、近くに住んでいる人の写真の方が有効なものが多いということがわかる。4月の写真については、13時台以降に撮影された写真が不要と選別されている。これらについては、これ以上の解釈は難しい。

①-5 ルールの構築：ラーメン

それぞれの説明変数候補における、有効写真と不要写真の比較を、表 15 と表 16、図 56 から図 59 に示した。

表 15 属性値の記録率（ラーメン）

属性値		有効	不要
Flicker 側 の メ タ デ ー タ	タイトル	99.8%	100.0%
	タグ	86.2%	92.7%
	撮影日時	100.0%	100.0%
	居住地	54.9%	62.7%
Exif デ ー タ	ファイル ソース	17.6%	39.0%
	機器メーカー	54.4%	81.9%
	シャッター スピード	53.7%	81.9%

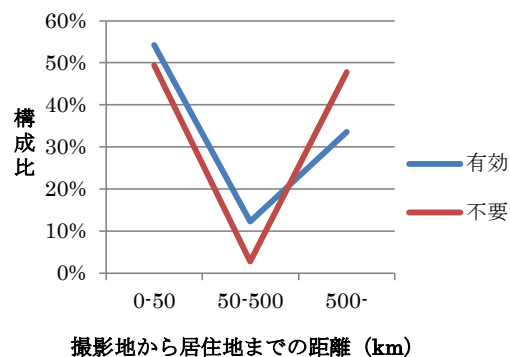


図 56 撮影地から居住地までの距離(ラーメン)

表 16 機器メーカーごとの不要写真率(ラーメン)

	写真枚数	不要写真枚数	不要写真率
Apple	259	92	35.5%
その他	292	143	49.0%
総計	551	235	42.6%

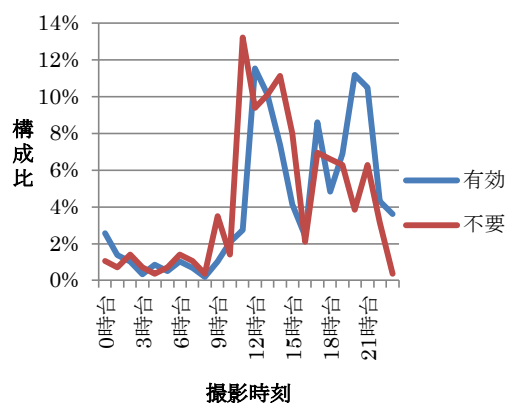


図 57 撮影時刻（ラーメン）

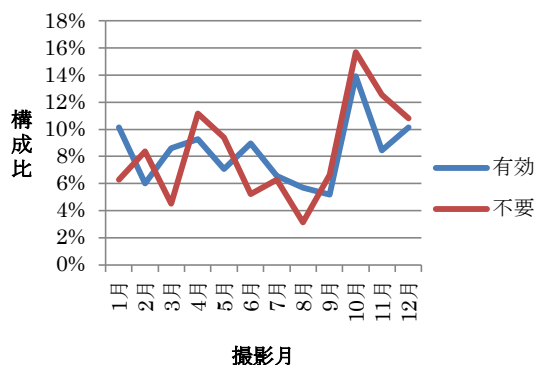


図 58 撮影月（ラーメン）

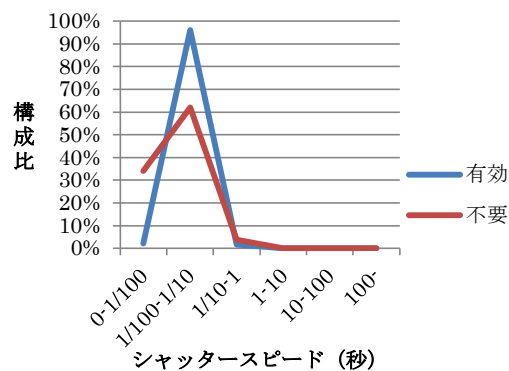


図 59 シャッタースピード（ラーメン）

決定木

傾向に差が見られたのは、③ファイルソースの有無、④居住地登録の有無、⑤居住地と撮影地の距離、⑦撮影時刻、⑨シャッタースピードの、5つの説明変数である。この5つの説明変数を用いて得られた決定木を図60に示し、その解釈を行った。

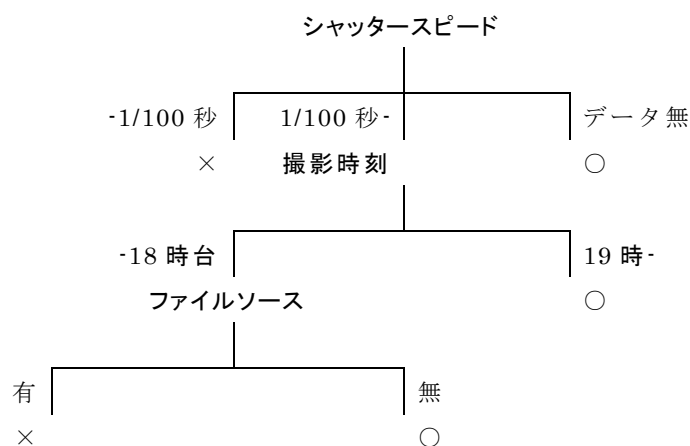


図 60 ラーメン写真選別の決定木

第一の選別の鍵は、シャッタースピードである。シャッタースピードが 1/100 秒未満で速いものは、不要と選別されている。ラーメンは、店内で撮影されることがほとんどなので、そこまで明るい状況での撮影はなく、シャッタースピードが速くなるということが少ないと考えられる。また、1/100 秒以上の遅いもののうち、19 時以前にデジタルカメラで撮影されたもの以外は、有効と選別されている。ラーメンをわざわざデジタルカメラで撮影することは少ないのでは、というようなことが考えられる。

①-6 ルール構築：東京スカイツリー

それぞれの説明変数候補における、有効写真と不要写真の比較を、表 17 と表 18、図 61 から図 65 に示した。

表 17 属性値の記録率（東京スカイツリー）

	属性値	有効	不要
Flickr 側の メタ デー タ	タイトル	99.2%	90.9%
	タグ	92.4%	99.1%
	撮影日時	100.0%	100.0%
	居住地	76.5%	74.5%
Exif デー タ	ファイル ソース	54.4%	24.7%
	機器メーカー	87.0%	88.6%
	シャッター スピード	86.3%	86.7%

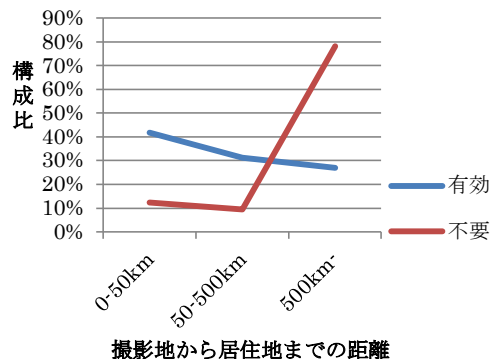


図 61 撮影地から居住地までの距離（東京スカイツリー）

表 18 機器メーカーごとの不要写真率（東京スカイツリー）

	写真枚数	不要写真枚数	不要写真率
Apple	446	373	83.6%
その他	2089	935	44.8%
総計	2535	1308	51.6%

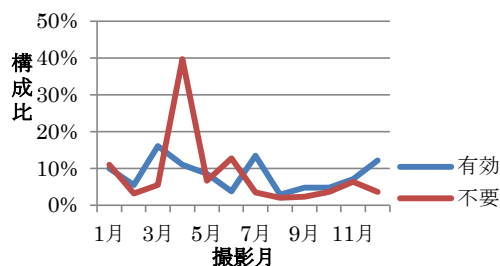


図 63 撮影月（東京スカイツリー）

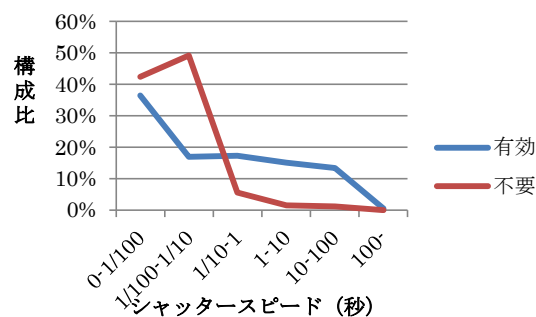


図 64 シャッタースピード（東京スカイツリー）

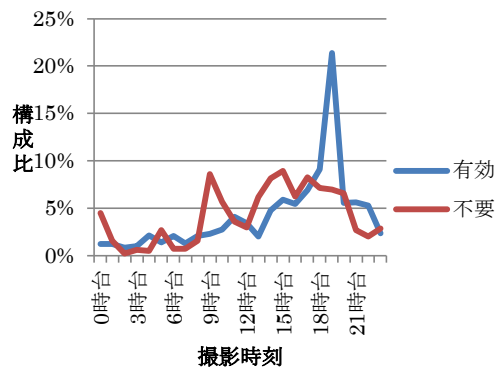


図 62 撮影時刻（東京スカイツリー）

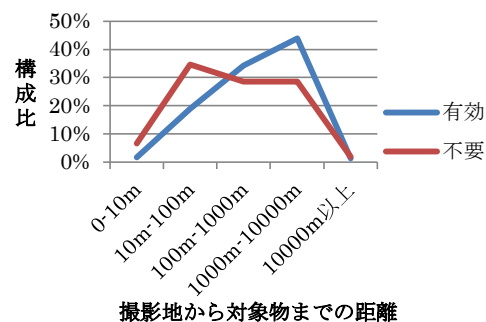


図 65 撮影地から対象地までの距離（東京スカイツリー）

決定木

傾向に差が見られたのは、①タイトル有無、②タグ有無、③ファイルソース有無、⑤撮影地から居住地までの距離、⑥機器メーカー、⑦撮影時刻、⑨シャッタースピード、⑩撮影地から対象物までの距離、の8つの変数である。この8つの説明変数を用いて得られた決定木を図66に示し、その解釈を行った。

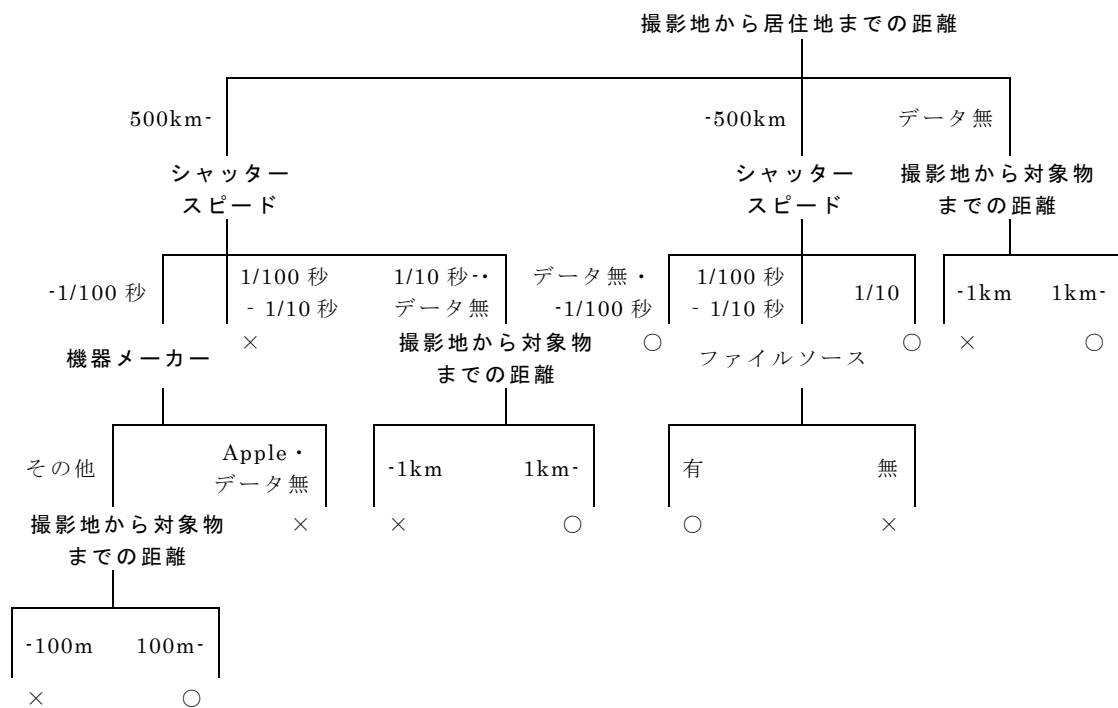


図 66 東京スカイツリー写真選別の決定木

第一の選別の鍵は、撮影地から居住地までの距離である。大まかに見ると、500km 圏外からの旅行者の写真は、不要なものの割合が多く、500km 圏内の旅行者の写真は有効なものの割合が多いということになる。遠方から来ている人の写真は、「スカイツリー」というタイトルやタグがついていても、関係のない写真であることが多いということである。もしくは、対象物までの距離が近いものが不要と選別されていることから、スカイツリー内部での写真を多く投稿しているのではないかと考えられる。

②ルール適用

構築したルールを、東京の各テーマの全データに適用した結果が表 19、大阪のデータに適用した結果が表 20 である。

表 19 構築したルールの、東京におけるデータへの適用結果

	適用前有効率	適用後有効率	適用後 有効写真残存率
夜（狭い定義）	61.4%	77.2%	60.7%
夜（広い定義）	80.4%	93.1%	58.2%
猫	67.4%	91.2%	77.4%
桜	86.0%	92.6%	75.9%
ラーメン	66.9%	77.2%	90.7%
東京スカイツリー	48.8%	80.2%	76.4%

表 20 構築したルールの、大阪におけるデータへの適用結果

	適用前有効率	適用後有効率	適用後 有効写真残存率
夜（狭い定義）	35.0%	51.8%	64.7%
夜（広い定義）	46.4%	93.2%	44.2%
猫	31.8%	33.3%	2.9%
桜	91.3%	94.9%	27.2%
ラーメン	53.7%	71.2%	89.3%

東京スカイツリーを除く 4 テーマについては、東京の全データに適用した場合はもちろんのこと、大阪のデータに適用した場合も、全てのテーマにおいて有効写真率が上がっている。ただし、夜（広い定義）、猫、桜に関しては、大阪のデータへの適用後の有効写真残存率が 50%を下回っている。特に、猫に関してはその減少が著しい。もしも手動で判別した場合は、判別後の有効写真率、有効写真残存率はいずれも 100%となり、この割合をいかに高く保てるかが、有効な自動選別ルールの構築に求められていることである。

テーマ別に見ていくと、猫は、大阪のデータに適用した場合、適用前は 34 枚だった有効写真が 1 枚になっている。わずかに有効写真率が上がっているとはいえ、これでは有効なルールとは言えない。そもそも適用前の有効写真枚数

も、他のテーマと比べて極端に少ない。東京の場合は、一部の人が大量に同じ猫の写真を投稿しているために合計枚数が多くなっているが、本来は撮影機会の少ないテーマなのかもしれない。また、一部の投稿者による大量投稿のために、その一部の投稿者の情報がルール構築の際の説明変数に大きく影響し、それが有効なルールを構築できなかった一因になったとも考えられる。

夜は、広い定義の方では特に大阪のデータへのルール適用後の有効写真率が上がっている。夜（広い定義）というのは、「夜の可能性が高いと客観的に判断できる写真」を有効写真に含めたものである。得られた決定木を見ると、選別の一番の鍵は、「シャッタースピード」である。シャッタースピードの長い写真、つまり多くの光を必要とする暗い場所での写真を有効と判別するのがこの決定木の一段階目である。目視で「夜である」、もしくは「夜の可能性が高い」と判断できる写真は、シャッタースピードのデータを用いて選別できることがわかった。投稿者の特性ではなく、夜というテーマの特性がルールの鍵になったために、他地域のデータに適用した場合にも有効なルールが構築できたのだと考えられる。

ラーメンは、東京においても大阪においても、ルール適用後の有効写真残存率が高い。第一の選別の鍵は、夜同様、シャッタースピードである。ラーメンの写真というのは、ほとんどがラーメン店内において、近距離でラーメンの入ったどんぶりが撮影されたものである。つまり、撮影する人や地域が変わっても、写真の撮影される状況には変わりがなく、テーマの特性に合ったルールを構築することができたと考えられる。有効写真に共通点があるので、残存率が高い。しかし、有効写真率は大阪のデータへの適用で約 70%にとどまっているので、不要写真を排除しきれていないということであり、ルールの構築には課題が残る。

桜は、東京においても大阪においても、適用前から有効写真率の高いテーマであるが、ルール適用後はさらに有効写真率が向上している。選別の一番の鍵である撮影月という説明変数は、テーマの特性に合っている。ただし、大阪の

データでは、写真枚数が大幅に減少している。東京のデータにおいて、このテーマはジニ係数が高めであったため、一部の投稿者の情報がルール構築の際に影響した可能性がある。また、東京の桜の有効写真を見てみると、一般的に見頃な3月～4月以外にも、投稿がある。晩冬から早春に咲く寒桜などであると思われる。本研究において構築されたルールでは、そのような、CGM データを用いてこそ得られた一般的ではない情報が消えることになってしまうので、作成したいマップの特徴を考えて、説明変数を選ぶことが必要である。

5.3 調査2のまとめ

以上の調査から、有効なルールを構築するためには、テーマの特性（撮影対象の特徴）に合った説明変数を用いることが必要であるということがわかった。猫のようにジニ係数の高いテーマ、つまり投稿者ごとの投稿枚数の偏りが大きいテーマについては、そのことに特に注意してルールの構築や、ルール作成時に使う元の写真のある程度の選別を行う必要がある。

今回新たに用意したフィルタは、撮影月、シャッタースピード、撮影地と対象地の距離（東京スカイツリーに対してのみ）だが、いずれも自動選別ルールに組み込まれ、桜においては撮影月が、夜（狭い定義・広い定義）、ラーメンにおいてはシャッタースピードが、第一の選別の鍵となっている。まだ、今回考察に絡めていない Exif データがさらに数多くあるので、説明変数は増やせるはずである。例えば、シャッタースピードが夜の写真の選別に使えたのなら、おそらく絞り値も使えるだろう。つまり、この決定木は、改良の余地が残されている。

VI. まとめと課題

本研究ではまず、地図化を行いたいテーマごとに、抽出されたデータを分析し考察することで地図化適性に影響する要因を整理した。その結果、キーワード検索によって写真を抽出するだけでは、そのテーマのマップ作れるとは言えないということがわかった。なぜなら、様々な影響により、意図する写真が得られない、もしくは意図しない写真まで抽出することになるためである。また、データ数が多いテーマに対しては自動選別ルールの構築を行い、有効なルール構築に必要な説明変数の性質を明らかにすることで、テーマ別観光マップの作成手法の一部を示した。

本研究における方法では、あくまで写真に収められるものしか地図化することができない。テーマによって、撮影機会や撮影難度が異なり、関心があるからと言って写真を撮るとは限らない。タグやタイトルにキーワードが含まれているものしか抽出できない。さらに、表記ゆれや、タグ・タイトルの付け方によっては、抽出したい写真を得られなかったり、余計な写真を抽出してしまったりすることもあり得る。また、人の目で判別できるものしか地図化できない。つまり桜や猫など、一般的なものの分布を地図化することはできるが、「可愛い猫」「満開の桜」というような、程度の差や個人の主観に判断をゆだねる種類のものは、目視で判別することが難しく、地図化に向かない。また、雨などは、たとえ写真に写っていたとしても、それをマッピングしたところで、それがどのような意味を持つのかは、解釈が難しいところである。地図化にあたっての課題としては、投稿者ごとの投稿写真枚数の偏りの考慮や、時期を細かく区切った表示、他テーマや他期間との比較によってポテンシャルの高さを示す方法などが挙げられ、表示ツールの改良の余地が大いにあることが示された。

これらを受け今後は、地図化するテーマの選定、地図化の際の表現の工夫、有効な自動選別ルールの構築、という面において残されている課題に取り組み、テーマ別観光マップの実用化を目指したい。

謝辞

本論文を書くに当たり、終始ご指導ご助言を頂きました、指導教官及び主査である首都大学東京大学院都市環境科学研究科観光科学域倉田陽平准教授、及び副査を務めて頂きました本保芳明教授、菊地俊夫教授には、深く感謝いたします。また、ゼミにおいて観光政策・政策領域の皆様より沢山の意見をいただいたこと、感謝いたします。誠にありがとうございました。

(注)

- 1) 写真コミュニティサイト。言語は英語。<https://www.flickr.com/>
- 2) 写真コミュニティサイト。言語は日本語。<http://photozou.jp/>
- 3) 株式会社ゼロインが運営する、訪日外国人向けの、ハウツーサイト。
<http://www.deepjapan.org/>
- 4) 気象庁が公開している、過去の「東京」における、10分毎の降水データと、写真の撮影時刻を紐づけて、雨天時に撮影されたであろう写真を抽出した。

参考文献

- 青野由季・倉田陽平(2014) 雨の日の観光の可能性 ―若者に対する意識調査が示唆するもの―. 観光情報学会第11回全国大会, 50-51, 柏, 2014年6月.
- 王佳な・野田雅文・高橋友和・出口大輔・井手一郎・村瀬洋(2011) Web上の大量の写真を画像分類して提示する観光マップの提案. 電子情報通信学会技術研究報告. IE, 画像工学 110(456), 153-158, 2011-02-28
- 奥山幸也・柳井啓司(2011) 写真撮影の位置軌跡を利用した旅行支援システム. DEIM Forum, F7-6
- 倉田 陽平・杉本 興運・矢部 直人(2010) あえて案内しない着地型観光案内ー観光関心点データの抽出と活用. 第19回地理情報システム学会学術大会, 地理情報システム学会講演論文集 19, CD-ROM, 京都, 2010年10月.
- 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観光情報サービス. 第20回地理情報システム学会学術大会, 地理情報システム学会講演論文集 20, CD-ROM, 鹿児島, 2011年10月.
- 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回研究発表大会, 大阪府泉南郡熊取町, 2012年12月.
- 倉田陽平(2013) 観光ポテンシャルマップの信頼性向上に向けてーソースとなる投稿写真データの自動選別ルール of 構築ー. 第22回地理情報システム学会学術大会, 東京, 2013年10月, CD-ROM
- 倉島 健・岩田 具治・入江 豪・藤村 考(2010) 写真共有サイトにおけるジオタグ情報を利用したトラベルルート推薦. 電子情報通信学会技術研究報告. LOIS, ライフインテリジェンスとオフィス情報システム 109(450), 55-60, 2010-02-25
- 白柳かさね(2011) 宮城県石巻市田代島における猫観光の地域的背景ー人間と猫の関わりに着目してー. 首都大学東京都市環境科学研究科観光科学域 2010年度修士論文概要集, 33-40

杉本興運（2011）デジタルカメラを活用した写真撮影調査の検討：社会動向，
調査時の意識，撮影枚数の個人差について．首都大学東京 大学院都市環境
科学研究科 観光科学域 観光科学研究 4,89-98.

長尾光悦（2012）CGM からの観光情報の獲得．情報処理学会第 74 回全国大
会講演論文集 2012(1), 505-507, 2012-03-06

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.

Chen, W. C., Battestini, A., Gelfand, N., & Setlur, V. 2009. Visual
summaries of popular landmarks from community photo collections.
ACM Multimedia Conference: 789-792.

De Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R.,
& Yu, C. 2010. Automatic construction of travel itineraries using social
breadcrumbs. *ACM Conference on Hypertext and Hypermedia*: 35-44.

Lu, X., Wang, C., Yang, J., Pang, Y., and Zhang, L. 2010. Photo2Trip:
Generating Travel Routes from Geo-Tagged Photos for Trip Planning.
Intl. Conf. on Multimedia, 143-152.

気象庁 <http://www.jma.go.jp/jma/index.html>（アクセス日 2015年1月5日）

東京 Zoo ネット <http://www.tokyo-zoo.net/>（アクセス日 2015年1月5日）

マイナビニュース <http://news.mynavi.jp/>（アクセス日 2015年1月5日）

日テレ NEWS24 <http://www.news24.jp/index.html>（アクセス日 2015年
1月5日）