

写真共有サイトを用いた穴場スポットの抽出

西脇 達也[†] 北山 大輔[†]

[†] 工学院大学情報学部コンピュータ科学科 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: tj111078@ns.kogakuin.ac.jp, tikitayama@cc.kogakuin.ac.jp

あらまし 観光スポットには、景観がよいものの、知名度が低い地点が存在する。我々は、このような、知名度が低い他の地点と比べて同等以上の満足が得られる地点を穴場スポットと定義した。本研究では、画像共有サイト flickr より、タグおよび位置情報が付与された写真データを抽出しクラスタリングを行うことで、ある地点周辺の穴場スポットの候補を得る。そして、クラスタごとにお気に入り数と写真数を用いて穴場スポット度を算出することで穴場スポットを得る手法を提案する。

キーワード 観光情報, 位置情報, DBSCAN, GIS

1. はじめに

観光スポットには、景観を眺めるために適した地点が複数存在することがある。たとえば、東京タワーについて考えてみると、東京タワーと夕日を眺めることができる地点と、東京タワーと富士山を眺めうことができる地点などがあげられる。その中には、知名度が低くまた、他の地点と同等以上の満足が得られる地点が存在する場合がある。地点ごとに異なる景観を眺めることができ、それぞれに魅力があると考えられる。本研究では、このような地点を景観地点と呼ぶ。景観地点には、人が多く集める有名な地点と、そうではない知名度が低い地点がある。有名な地点では、多くの人が集まるため、魅力的な景観地点であるといえる。一方、知名度が低い地点にも魅力的な景観を見ることができる地点が存在している場合がある。しかし、観光スポットの情報を得る際には、人気な景観地点が多く、知名度が低い景観地点を探すことは困難である。旅行前の計画時によく用いられる、ガイドブックや観光のポータルサイトは、主として、有名なスポットしか掲載されておらず、再訪時の計画を立てる際に参考にならない場合がある。また、旅行中においても、空き時間ができ訪問場所を探す場合にも、ガイドブック等には、載らないが、有用な地点が見つげられることが望ましい。

そこで我々は、このような知名度が低い、他の景観地点と同等以上の満足が得られるような景観地点を穴場スポットと定義し、穴場スポットを抽出する手法を提案する。本研究の概念を図1に示す。この図では、観光スポットに対し3つの景観地点が存在する。左側2つの地点は、多くの人に知られており、評価も良い地点である。右側の地点については、人数は少ないが他の地点と同等程度の魅力のある地点であると考えられる。本研究では、右側の地点を穴場スポットと定義し、抽出する。その際に、穴場スポット度を算出し、値が高い地点を穴場スポット候補として出力する。

近年、多くのユーザにより Web 上に大量の写真がアップロー

ドされている。そして、flickr^(注1)を代表とする、写真共有サイトの登場により、写真を他人と共有することができる。flickr は、2012年に共有写真数が70億枚に達したと発表しており、膨大な量の写真が存在する。位置情報付きの写真は、旅行時に撮影されたものが多く、写真閲覧者は、撮影場所の景観情報を知ることができる。そのため、観光情報を得るための重要なツールの一つと考え、穴場スポット度の算出のデータとして用いる。本研究では、多くの写真がアップロードされている、flickrを使用する。flickrにアップロードされている写真には、位置情報、撮影日やお気に入り数などのタグが付与されている。

本稿では、2章で関連研究について述べ、3章で穴場スポットを推定に必要なクラスタリングと穴場スポット度の算出方法について述べる。4章でクラスタリングの閾値決定のための予備実験を行い、5章で評価実験について述べる。

2. 関連研究

Web上の位置情報付きデータを利用した観光情報システムに関する研究は、多数存在する。白井らの研究[1]では、位置情報付き写真を用いて多量の写真が撮影された複数の地点を可視化、また、撮影方向を利用しランドマークの形状の抽出手法を提案している。熊野らの研究[2]では、観光スポットには、時期により人々が集まる人数が異なることに注目し、撮影スポットの旬のシーズンを抽出する研究を行った。旬のシーズンを抽出する際に、バースト性に着目した指標を提案している。この指標は、その地域において注目するシーズンが短いまたは、シーズン期間内に撮影者数が多ければ、大きくなるように定義されており、この値によりシーズンを特定する。王らの研究[3]では、ジオタグ付き写真をクラスタリングし、各クラスタが風景を表していると考え、そのアイコンをマップ上に表示する手法を提案している。これらの研究では、写真が密集している地点やバースト性といった、写真数が多い場所へ着目しているが、本研究では、比較的照片数が少ないつまり、あまり知られていない場所に重みをおく点で異なる。

(注1) : <https://www.flickr.com/>

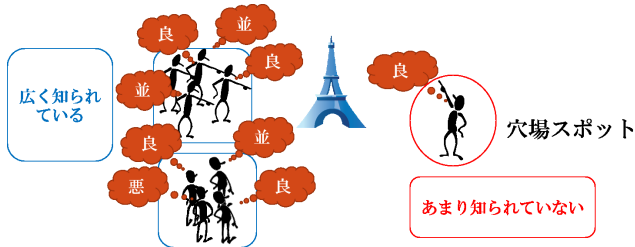


図1 本研究の概念図

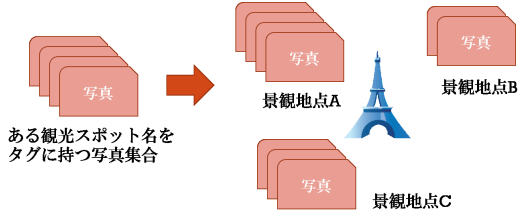


図2 本手法の概要

奥山らの研究[4]では、写真に付与された位置情報を連続して考慮することにより、移動軌跡と考える。この移動軌跡を、連結させることにより、新しい観光ルートをユーザに推薦する手法を提案している。奥らの研究[5]では、観光スポット領域内のツイートを基に、観光スポットの特徴を抽出する手法を提案している。今井らの研究[6]では、位置情報付きツイートを使用し、ユーザの日常、非日常行動範囲を推定する手法を提案している。上野らの研究[7]では、観光スポットの時間的特徴抽出を提案している。観光スポット領域を特定し、スポット領域内のツイートを発信時刻に基づいて魅力度を算出している。奥らの研究[8]では、対象地域内の文章に対して形態素解析を行い、IDFと共起度を用いて地域限定語句の抽出を提案している。

3. 穴場スポットの推定

1つの観光スポットに対して、複数の景観地点があると仮定する。本研究では、このような景観地点を、観光スポットにおける穴場スポットと定義する。

知名度に関して、多くの人が知っているという定義も考えられるが、本研究においては、地点に対する知名度としては、その地点に対する滞在数が知名度を表すと考えた。滞在数としては、TwitterやFoursquareなどのジオタグ付きコンテンツが利用可能である。本研究では、写真を撮影する行為は、その地点の評価であり、かつ、その地点への滞在を表すと考え、写真共有サイトを用いる。本手法の概要を図2に示し、以下にまとめる。

初めに、flickrAPIを利用し、特定の観光スポットをタグにもつ写真のデータを取得する。取得した写真データの中で、位置情報が付与された写真のみを使用する。次に、写真に付与された位置情報に基づいてクラスタリングする。この各クラスが、その観光スポットに対する景観地点を表している。最後に得られたクラスごとに人気数と写真数を用いて穴場スポット度を算出する。これにより、滞在数に対する満足度が高い地点を抽出する。

3.1 クラスタリング

写真に付与された位置情報に基づいてクラスタリングを行い、1つの観光スポットに多数存在する、景観地点を発見する。景観地点では、多くの写真が撮影されると考えられるため、写真が密集している。そのため、クラスタリング手法は、DBSCANを用いる。DBSCAN[9][10]は、クラスターの密度を基準にクラスタリングを行うため、高密度なクラスターのみを抽出することが可能となる。DBSCANは、クラスター間の距離の閾値 Eps とクラスターのデータ数の閾値 $MinP$ との2つの閾値をもつ。ある点 x から、距離 Eps 内にある点集合を近傍 $N_{Eps}(x)$ と定義し、以下の接続関係を満たすとき、同じクラスターに分類する。

- (1) $y \in N_{Eps}(x)$
- (2) $|N_{Eps}(x)| \geq MinP$

ただし、 $N_{Eps}(x) = \{y \in X | D(x, y) \leq Eps\}$ である。つまり、ある座標から半径 Eps 内に $MinP$ 以上の座標集合が存在するならば、同じクラスターに分類する。観光スポットによって、景観地点の分布は異なる。特に遠くから見えるスカイツリーのような観光スポットと、清水寺のような、内部が主な景観地点になる観光スポットでは、距離の閾値が異なると考えられる。そのため、距離の閾値が Eps は、写真データの位置の分布によって決定する必要がある。本研究では、写真データの平均地点からの距離の平均を基準に Eps を決定する。具体的には以下の手順で算出し、図3に表す。

- (1) 全座標データ $n = \{n_1, n_2, n_3, \dots, n_n\}$ の平均座標 t を算出し、全座標から t までの距離 $u = \{u_1, u_2, u_3, \dots, u_n\}$ を算出する。
 - (2) 集合 u 中の距離の上位5%個座標データをから外れ値として n から除外する。
 - (3) 集合 n から再度、平均座標 t' を算出し、 n の各要素から t' までの平均距離を ave とする。
- Eps を以下の式で定義する。

$$Eps = \frac{ave}{m} \quad (1)$$

ここで m は、平均距離をもとにした Eps の値を決定するための定数パラメータであり、予備実験によりもとめる。

3.2 穴場スポット度の算出

穴場スポット度の値が高いクラスターを穴場スポットとして抽出する。穴場スポットとして重要なポイントは、知名度が低いこと、他の景観地点と同等以上の満足度が得られることである。知名度が低い景観地点は、撮影者数が少ないと考えられる。また、景観地点の満足度は、景観地点内の各写真に付与されているお気に入り数で定義した。写真に付与されるお気に入りには、写真閲覧者が自らの意思で付与することができ、その写真の景観に何らかの魅力を感じて付けていると考えられるからである。よって、クラスター C_i の穴場スポット度 GUP_i は、クラスター内の撮影者数 $|C_i|$ 、クラスター内の写真の総お気に入り数 F_i であるとき、次式で表される。

$$GUP_i = \frac{\log_{10} F_i}{|C_i|} \quad (2)$$

すなわち、クラスター内の撮影者数が少なく、お気に入り数が大

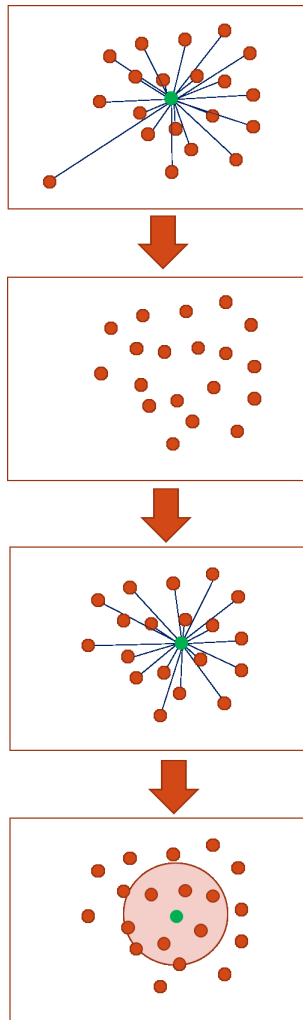


図3 分布に基づく Eps のパラメータ決定

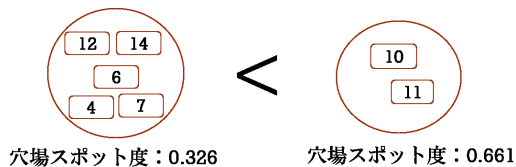


図4 穴場スポット度算出例

きければ、より穴場スポットらしいと判断できる。例を図4に示す。四角は写真を表しており、数字は、その写真に付与されているお気に入り数を表している。この例では、各写真は異なるユーザが撮影したとする。

4. 予備実験

位置情報に基づくクラスタリングを実装する際、DBSCANを用いる。このとき、距離の閾値 Eps とデータ数の閾値 $MinP$ の値によりクラスタリング結果が異なる。よって、2つの閾値を定めなければならない。そのため、予備実験を行い閾値を決定する。

これら2つの閾値を設定するにあたり、クラスタリング結果の正解データを用意した。我々が用意した正解データは、六義園

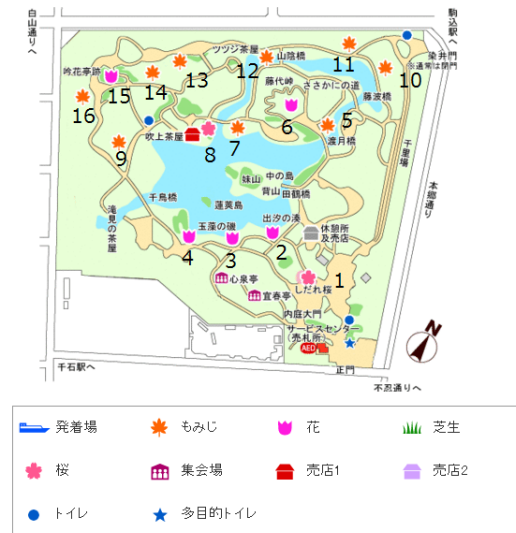


図5 六義園における正解データ



図6 クラスタリング結果(六義園)

園^(注2)であり、図5に示す。今回は写真が多く撮影されるであろう「もみじ」、「花」、「桜」3種類、計16カ所を景観地点と考え実験を行う。flickr に対し、「六義園」のタグを持つものを検索し、位置情報を持つ989件のデータをクラスタリングした。そして、正解データをもっとも網羅できるようにクラスタが生成される m と $MinP$ を求める

クラスタリングの結果を図6、詳細を表1に示す。図6の○は、正解地点にクラスタが存在する場所を表している。表1は、 m と $MinP$ の値を変更したとき、16カ所の正解地点をどのくらい含むことができたかそして、正解地点にクラスタが存在しない数の変化を表している。最終的に14カ所の正解を含み、外れ正解数が1つの $m = 9$, $MinP = 2$ に定まった。

5. 評価実験

5.1 目的と内容

撮影者数が2人以上の地点かつ、穴場スポット度が高い地点について評価を行う。2人以上とした理由は、1人だと偶然撮

(注2) : 庭園へ行こう: <http://teien.tokyo-park.or.jp/contents/index031.html>

表 2 穴場地点の候補

順位	座標	穴場スポット度	クラスタ番号	地名	穴場判断
1	35.7323138,138.3716424	0.89619	15	韭崎市 (山梨)	○
2	35.4507503,139.8969736	0.75925	33	木更津市 (千葉)	○
3	36.2852035,137.638469	0.63937	16	松本市 (長野)	×
4	35.6823255,138.3798578	0.58804	13	韭崎市 (山梨)	○
5	35.14231667,139.6311417	0.54108	9	三浦市 (神奈川)	○

表 1 予備実験結果の詳細

m	MinP	クラスタ数	正解を含むクラスタ数	外れ正解数
7	2	31	12	1
7	3	28	12	1
7	4	22	11	2
7	5	17	10	3
8	2	39	13	1
8	3	31	12	2
8	4	25	12	3
8	5	17	10	4
9	2	40	14	1
9	3	34	13	2
9	4	24	12	3
9	5	18	10	6
10	2	45	14	2
10	3	37	12	4
10	4	28	12	4
10	5	23	10	4

影した地点が抽出された可能性があるからである。評価方法は、Web 検索を行い、穴場としていわれている地点であるかどうか調査する。実際に穴場であるかどうかの判断は非常に難しく、現地に訪れた主観的な評価が望ましいと考えられるが、本稿では、簡易的に Web 上の記述があるものについて評価する。flickrAPI を使用し、ある観光スポットをタグにもつ写真について検索を行い、位置情報が付与された写真データだけを使用する。実験ではまず、得られたデータに対しクラスタリングを行う。そして、分かれた各クラスタの中心に存在する地名を取得する。ここで、地名は、オブジェクト名が得られる場合は、オブジェクト名、得られな場合は、住所の市区町村を用いる。最後に、Web 検索により「地名 AND 観光スポット名」で得られるページにより判断する。本稿では、上位に出現するサイト 10 件中 2 件以上のサイトで、地名に対する観光スポットへの肯定的な意味の記述がされていればその地名は、その観光スポットに対して穴場スポットであると判断する。閾値は、六義園で得られた $m = 9$, $MinP = 2$ を使用する。観光スポットは、「富士山」で実験を行い、上位 5 つについて考察を行う。

5.2 結果と考察

表 2 は、穴場スポット度が上位 5 つのクラスタの地名と Web サイトでの判断結果を表す。また、図 8 に肯定的な記述の一部を、図 7 にクラスタリング結果を示す。図 7 の赤い円は、取得した上位 5 つの地点を表している。韭崎市が 2 つあるが、異なるクラスタで同じ市内に存在するからである。

上位 5 つの地点に関して、松本市以外で Web サイトで図 8

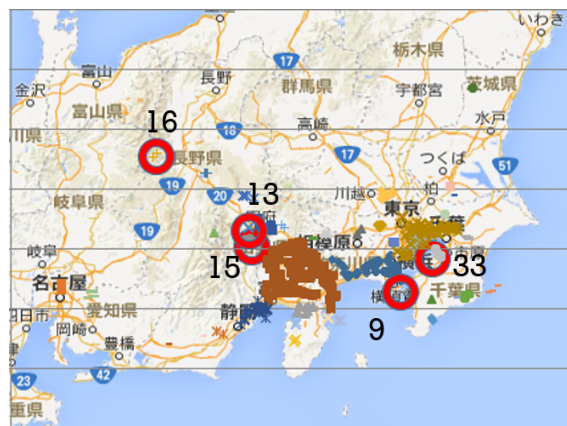


図 7 富士山のクラスタリング結果

夕焼けの時や、空気の澄んだ冬の朝の富士山は絶景です。映画「木更津キャッツアイ」で、恋人の橋として全国的に有名になりました。

橋からのながめは「ちば眺望100景」にも選ばれています。

数分で最初の群生地&絶景ポイントに到着します。富士山がどーんと見え、甲府盆地一体が見降ろせます。前回来た時は雲に隠れて殆ど見えませんでしたので初めて遠景を見たのですが、まさかこんなに絶景だったとは思いませんでした(びっくりでした)(笑)

大な夜景が楽しめます。山梨県でもトップクラスの夜景が楽しめるということもあり、夕暮れ時は数名の写真家が訪れていました。

図 8 Web から得られた肯定的な記述

のような肯定的な記述を確認することができ、魅力的な景色を眺めることができる地点を抽出することができた。また、富士山頂上と太陽が重なる現象「ダイヤモンド富士」を見ることができ、有用性を示すことができた。

6. ま と め

観光スポットに存在する景観地点のうち、知名度が低い但他的景観地点と同等以上の満足を得ることができる景観地点を穴場スポットと定義した。そして、撮影者数とお気に入り数を用いて穴場スポット度を算出し、穴場スポットを発見する手法を提案した。「富士山」をタグに持つ写真について実験し、穴場スポット度が上位 5 件に対して、魅力的な景観見ることができる地点を抽出することができた。今回は、Web を用いた評価実験を行ったが、得られた地点が評価されていることしか判定していない。本来は、知名度を含めた観点で評価する必要がある。そのため、アプリケーションとして公開し、実際に訪れた人にアンケート調査を実施するなどの評価実験が課題としてあげられる。また、Foursquare やレストラン検索を組み合わせること、提案手法を応用できると考えられるため観光オブジェクト

以外への適応も課題である。

謝 辞

本研究の一部は、平成 26 年度科研費若手研究 (B)(課題番号：24700098) によるものです。ここに記して謝意を表すものとします。

文 献

- [1] 白井 元浩, 廣田 雅春, 横山 昌平, 石川 博 ジオタグ付き写真を用いたホットスポットの分類とランドマークの形状抽出手法, DEIM Forum 2013 D8-1, (2013)
- [2] 熊野 雅仁, 岩渕 聡, 小関 基徳, 小野 景子, 木村 昌弘 集合知に基づいたポピュラー撮影スポットに関する旬シーズンの可視化, 芸術科学会論文誌 Vol.13, No.4, pp.218-228, (2013)
- [3] 王 佳な, 野田 雅文, 高橋 友和, 出口 大輔, 井手 一郎, 村瀬 洋 Web 上の大量の写真に対する画像分類による観光マップの作成, 情報処理学会論文誌 Vol.52, No.12, pp.3588-3592, (2011)
- [4] 奥山 幸也, 柳井 啓司 写真撮影の位置軌跡を利用した旅行支援システム, DEIM Forum 2011 F7-6, (2011)
- [5] 奥 健太, 橋本 拓也, 上野 弘毅, 服部 文夫 地理情報推薦のための観光スポットと位置情報付きユーザ生成コンテンツの対応付け手法の提案, DEIM Forum 2013 A3-2, (2013)
- [6] 今井 規善, 奥 健太, 服部 文夫 位置情報クラスタリングに基づく地理的ユーザプロファイリング手法, 情報処理学会第 75 回全国大会 1P-5, (2013)
- [7] 上野 弘毅, 奥 健太, 服部 文夫 位置情報付ユーザ生成コンテンツに基づくスポットの時間的特徴化の提案, 情報処理学会第 75 回全国大会 1P-6 (2013)
- [8] 奥 健太, 西崎 剛司, 服部 文夫 地域限定性スコアに基づく位置情報付きコンテンツからの地域限定語句の抽出, 情報処理学会論文誌 Vol.5, No.3, p97-116, (2012)
- [9] 神島 敏弘 データマイニング分野のクラスタリング手法 (2), 人工知能学会 18 巻 2 号, (2003)
- [10] Ester M, H-P Kriegel, Sander J, Xu x: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise (1996)