

土地利用分類の分布に基づく観光スポットの景観ラベルの推定

久保田 豪[†] 奥 健太[†]

[†] 龍谷大学理工学部 〒520-2194 滋賀県大津市瀬田大江町横谷 1-5
E-mail: [†]t150481@mail.ryukoku.ac.jp, ^{††}okukenta@rins.ryukoku.ac.jp

あらまし 入力された景観スポットに景観が類似するスポットを推薦するシステムを提案する．このような景観スポット推薦を実現するためには，スポットの景観的特徴化が課題となる．本研究では，国土数値情報として公開されている土地利用分類データに基づき，スポットの景観特徴化を行う．クラウドソーシングによりスポットに対し景観ラベルを付与し，それを教師ラベルとする．スポットの景観特徴と景観ラベルを基に学習モデルを生成し，そのモデルの性能を評価する．

キーワード 観光スポット推薦，景観特徴，土地利用分類

1. はじめに

2017 年の訪日外客数が 2,800 万人に達する^(注1)など観光を楽しむ人々は年々増大してきている．また，観光情報サイトの一つである TripAdvisor には，約 800 万件もの旅行情報（観光スポットやレストラン，ホテルなど）が登録されている^(注2)．観光者はこのような膨大な観光スポットの中から，訪れたい観光スポットを選択する必要がある．

このような背景の下，ユーザの嗜好に合った観光スポットを推薦する観光スポット推薦システム [1] [2] に関する研究が盛んに行われている．観光スポット推薦システムは，観光スポットの特徴とユーザの嗜好とのマッチングを行い，ユーザに観光スポットを推薦することが目的である．観光スポットの特徴としては，観光スポットのカテゴリ [3] や観光スポットに付与されたタグ [4] などが挙げられる．

我々は，観光スポットの特徴として景観特徴に着目し，景観特徴に基づく観光スポット推薦システムの実現を目指している．我々が景観特徴に着目する理由は，人々が観光を楽しむ際に，観光資源の景観という要素は無視できない重要な要素であると考えているからである．さらに，観光庁が平成 30 年度に講じようとする施策の一つとして，「景観の優れた観光資産の保全・活用による観光地の魅力向上」^(注3)を掲げていることから，観光資源における景観の重要性は裏付けられる．

景観特徴に基づく観光スポット推薦システムの実現に向けて，本研究では観光スポットの景観ラベル推定手法を提案する．本稿における景観ラベルとは，田園景観や山岳景観，森林景観など景観の種類を表すラベルを指す．本手法は，与えられた観光スポットがどのような景観ラベルに該当するかを推定する．

観光スポットの景観ラベルを推定するためには，観光スポットを記述する特徴量が必要となる．本研究では，観光スポット

の特徴量として土地利用分類データ^(注4)に着目する．土地利用分類データは，全国の土地利用の状況について，100m メッシュごとに利用区分を整備したものである．まず，与えられた観光スポット周辺の土地利用分類データの分布に基づき，観光スポットを特徴化する．また，あらかじめテストデータセットに含まれる観光スポットの景観ラベルを教師ラベルとして取得しておく．得られた観光スポットの特徴と教師ラベルを基に，景観ラベルを推定する学習モデルを生成する．

2. 関連研究

2.1 観光スポット推薦システム

観光スポット推薦システムの研究は多く行われている [1] [2]．観光スポット推薦システムは，観光スポットの特徴とユーザの嗜好とのマッチングを行い，ユーザに観光スポットを推薦することが目的である．

Gyorodi ら [4] は，ユーザに付与されたタグを用いた，内容に基づく観光スポット推薦システムを提案している．Lamsfus ら [5] は，オントロジーを用いて，ユーザおよび観光スポットの特徴を記述している．観光スポットの特徴を記述することができれば，内容に基づくフィルタリングを適用することが可能である．

これらの研究のように，内容に基づく観光スポット推薦システムにおいては，観光スポットを適切に特徴化する必要がある．これまでに，外部のデータに基づき観光スポットを特徴化する手法が提案されている [6] [7]．大田ら [6] は，観光スポットに関連付けられた位置情報付きツイートを基に感情語を抽出し，観光スポットを感情特徴化する手法を提案している．利田ら [7] は，Flickr に投稿された位置情報付き写真の投稿時刻を基に観光スポットを時間的特徴化する手法を提案している．

これらの研究 [6] [7] が，ツイートや投稿写真のユーザ生成コンテンツに基づく特徴化に着目しているのに対し，本研究では客観的なデータに基づく景観特徴に着目している．本研究は，観光スポット周辺の土地利用分類データの分布に基づき，観光

(注1) : <https://statistics.jnto.go.jp/>

(注2) : <https://tripadvisor.mediaroom.com/jp-about-us>

(注3) : 観光庁，平成 30 年版観光白書について（概要版），平成 30 年 6 月，<http://www.mlit.go.jp/common/001237304.pdf>

(注4) : <http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-L03-b.html>

スポットを景観特徴化する。

3. 定義

本章では、本稿に関連する各定義やデータ等について説明する。

観光資源および地域資源. 本稿では、清水寺や姫路城などの観光資源および白滝や比叡山などの地域資源を総称して資源とよぶ。資源 $r_i \in R$ は、その資源の名称 $r_i.name$ 、地理座標系の位置座標として経度 $r_i.x$ および緯度 $r_i.y$ をもつ。

地域メッシュ. 地域メッシュは、地域を緯度・経度に基づいてほぼ同一の大きさでメッシュ状に分割したものである。日本では標準地域メッシュ^(注5)が用いられている。標準地域メッシュは、その区分方法により第1次地域区画（1次メッシュ）、第2次地域区画（2次メッシュ）、第3次地域区画（基準地域メッシュ、3次メッシュ）などと定められている。

1次メッシュは、20万分の1地勢図の1図葉の区画を表し、1辺の長さが約80km（緯度差：40分、経度差：1度）となっている。2次メッシュは、1次メッシュを緯線方向および経線方向に8等分してできる区域であり、1辺の長さは約10km（緯度差：5分、経度差：7分30秒）である。3次メッシュは、2次メッシュを緯線方向および経線方向に10等分してできる区域であり、1辺の長さは約1km（緯度差：30秒、経度差：45秒）である。

各メッシュには、メッシュを識別するためのコードとして地域メッシュコードが割り振られている。例えば、北緯34度57分54秒、東経135度56分25秒に対応する3次メッシュコードは、5235-37-55となる。ここで、上4桁の5235は1次メッシュコードである。1次メッシュコードはメッシュの南西端の緯度 lat ・経度 lon を用いて次式で算出される：

$$(lat \times 1.5 \times 100) + (lon - 100) \quad (1)$$

次の2桁の37は2次メッシュコードである。1桁目が緯度方向の、2桁目が経度方向のコードを表す。これに1次メッシュコードを合わせて5235-37のように表す。下2桁の55は3次メッシュコードである。2次メッシュコードと同様に、1桁目が緯度方向の、2桁目が経度方向のコードを表す。これに1次・2次メッシュコードを合わせて5235-37-55のように表す。

土地利用分類データ. 土地利用分類データは、国土交通省により国土数値情報^(注6)として公開されている。土地利用分類データは、全国の土地利用の状況について、3次メッシュをさらに細分化した100mメッシュごとに、利用区分を整備したものである。利用区分として、表1に示す17種別が用意されている。

標高データ. 標高データ^(注7)は、国土交通省により国土数値情報として公開されている。標高データは、標高（平均、最高、最低）について、5次メッシュ（250mメッシュ）ごとに整備したものである。

(注5) : <http://www.stat.go.jp/data/mesh/>

(注6) : <http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-L03-b.html>

(注7) : <http://nlftp+.mlit.go.jp/ksj/gml/datalist/KsjTmplt-G04-d.html>

表1 土地利用分類における利用区分。

コード	種別
100	田
200	その他の農用地
500	森林
600	荒地
701	高層建物
702	工場
703	低層建物
704	低層建物（密集地）
901	道路
902	鉄道
1001	公共施設等用地
1002	空地
1003	公園・緑地
1100	河川地及び湖沼
1400	海浜
1500	海水域
1600	ゴルフ場

表2 景観ラベル。

景観ラベル	意味
rural	田園景観
mountain	山岳景観
volcano	火山景観
limestone	石灰岩景観
forest	森林景観
fishing-village	漁村景観
agricultural-village	農村景観
resort	リゾート景観
urban	都市景観
street	街路景観
road	道路景観
river	河川景観
port	港景観
lake	湖沼景観
coast	海岸景観
bridge	橋梁景観
dam	ダム景観

景観ラベル. 表2に示すラベルを景観ラベルとして扱う。これらの景観ラベルは、文献[8]や、国土数値情報の自然景観資源の大分類コード^(注8)を参考にして決定した。

4. 提案手法

提案手法の目的は、与えられた資源 $r_i \in R$ について、適切な景観ラベル $s_i \in S$ を自動的に付与することである。図1に提案手法の構成図を示す。提案手法は、大きく、(1) 資源への景観ラベルの付与（4.1節）、(2) 資源の地理的特徴化（4.2節）、(3) 景観ラベルの学習（4.3節）、(4) 景観ラベルの推定（4.4節）の四つの処理から構成される。以下、各処理について説明

(注8) : <http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-P19.html>

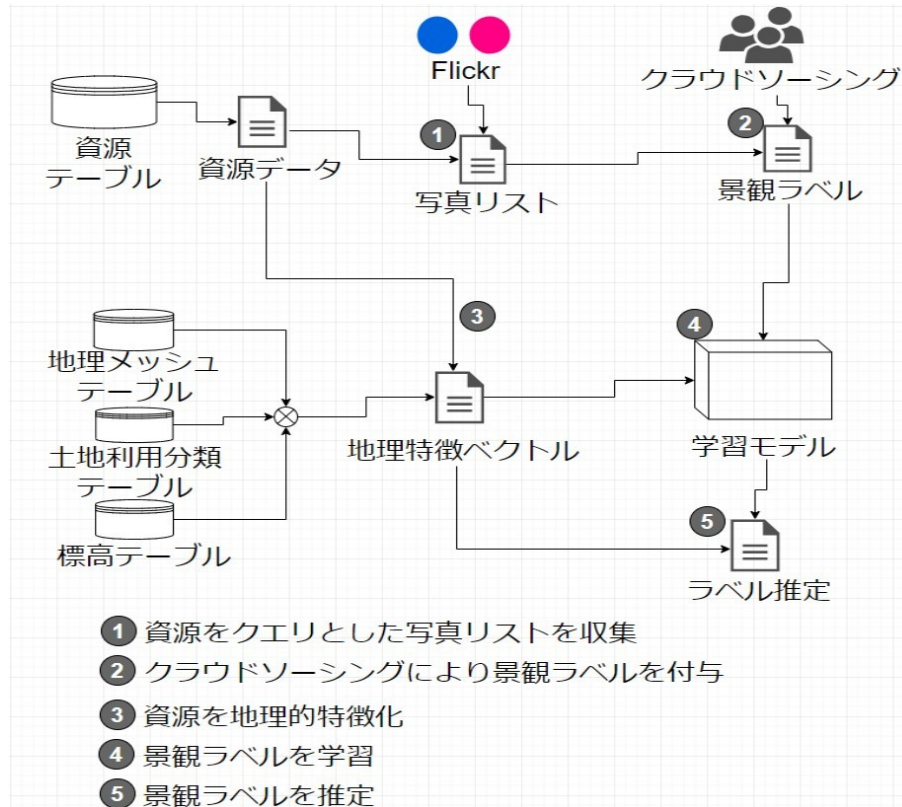


図 1 提案手法の構成図。

する。

4.1 資源への景観ラベルの付与

資源 $r_i \in R$ に対し、景観ラベル s_i を付与する。ここで付与する景観ラベルは学習時の教師ラベルとして用いる。

まず、写真共有サイトである Flickr^(注9) に対し、資源の名称 $r_i.name$ をクエリとして関連する写真を検索する。検索された写真リストを資源 r_i の写真リストとする。このとき、検索された写真数が 10 件未満の資源は以降の処理の対象から除外する。

残りの資源について、下記の手順により、クラウドソーシングを用いて資源 $r_i \in R$ に景観ラベル s_i を付与する：

(1) クラウドソーシングのワークに、資源 r_i の写真リストを提示する。

(2) ワークは、提示された写真リストに相応しいと思う景観ラベルを表 2 に示した候補の中から選択する。ここでは、複数選択も可能にしている。また、提示された写真リストのほとんどが景観を表すような写真でない場合は「景観ではない」、システムの不具合等により写真が表示されない場合は「写真が表示されない」を選択する。さらに、選択肢以外にも景観を表現するのに相応しい呼称があれば、「その他」を選択したうえで、「〇〇景観」という呼称で回答する。

資源 1 件あたりのワーク数は 5 名とする。資源 r_i に対して、5 名のうち 3 名以上が同一ラベルを付与したとき、そのラベルを資源 r_i の景観ラベル s_i とする。

4.2 資源の地理的特徴化

土地利用分類データおよび標高データに基づき、資源 r_i を

地理的特徴化する。地理的特徴化した資源 r_i を特徴ベクトル $g_i = (l_{0100}, l_{0200}, \dots, l_{1600}, e_{\max}, e_{\min}, e_{\text{avg}})$ で表現する。特徴ベクトル g_i は、20 次元ベクトルで表現され、1 次元目から 17 次元目までの要素 $l_{0100}, l_{0200}, \dots, l_{1600}$ が土地利用分類データに基づく要素、18 次元目から 20 次元目までの要素 $e_{\max}, e_{\min}, e_{\text{avg}}$ が標高データに基づく要素である。

まず、資源 r_i の経度・緯度 $(r_i.x, r_i.y)$ を基に、その資源が属す地域メッシュを取得する。この地域メッシュを基に、土地利用分類データに基づく要素と標高データに基づく要素をそれぞれ抽出する。ここで、地域メッシュの選び方としては 2 次メッシュまたは 3 次メッシュを考える。これらの選び方の違いについては ?? 章で検証する。

各データに基づく要素の抽出については、それぞれ下記の手順により行う。

a) 土地利用分類データに基づく要素の抽出。

資源 r_i が属す地域メッシュに含まれる土地利用分類の種別ごとの個数を対応する要素の値とする。例えば、メッシュ中に「田」の種別が 37 件含まれているとすると、田に対応する要素の値が $l_{0100} = 37$ となる。

b) 標高データに基づく要素の抽出。

資源 r_i が属す地域メッシュに含まれる標高データから、最高標高、最低標高、平均標高をそれぞれ取得する。それぞれ、最高標高を $e_{\max}$ に、最低標高を $e_{\min}$ に、平均標高を e_{avg} に格納する。

図 2 に、土地利用分類データおよび標高データに基づく要素の抽出例を示す。図の左側が土地利用分類の種別の分布を、

(注9) : <https://www.flickr.com/>

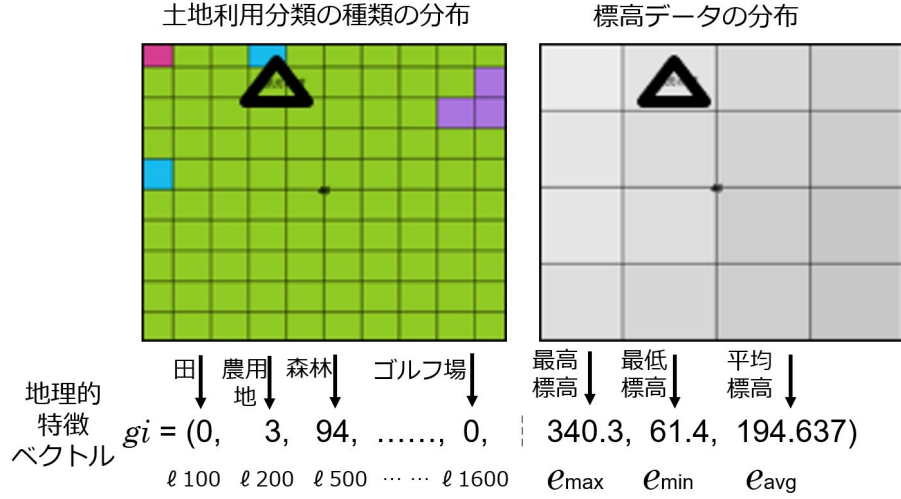


図 2 図の左側が土地利用分類の種類の分布を、右側が標高データの分布を示しており、どちらも同一の領域を示している。図中の三角がスポットの位置を表している。

右側が標高データの分布を示しており、どちらも同一の領域を示している。図中の三角がスポットの位置を表している。図の例では、土地利用分類は、例えば“田”が0件，“農用地”が3件，“森林”が94件が含まれている。また、標高データの中で、最高標高が340.3m、最低標高が61.4m、平均標高が194.637mであるということを示している。それぞれ得られた数値を各要素の値として、資源 r_i の特徴ベクトル g_i とする。

4.3 景観ラベルの学習

資源 r_i について得られた地理的特徴ベクトルと景観ラベルの対 (g_i, s_i) を学習データとし、機械学習により学習モデルを生成する。本研究では、機械学習手法として SVM を用いる。

4.4 景観ラベルの推定

生成した学習モデルを用い、未知の資源 r_k の景観ラベル $\hat{s}_k$ を推定する。ここで、資源 r_k の地理的特徴化は 4.2 節と同様である。

5. おわりに

本研究では、景観特徴に基づく観光スポット推薦システムの実現に向けて、観光スポットの景観ラベル推定手法を提案した。提案手法では、まず、与えられた観光スポット周辺の土地利用分類データの分布に基づき、観光スポットを特徴化する。また、あらかじめテストデータセットに含まれる観光スポットの景観ラベルを教師ラベルとして取得しておく。得られた観光スポットの特徴と教師ラベルを基に、景観ラベルを推定する学習モデルを生成する。

今後は、実データを用いた実験により、提案手法による景観

ラベルの推定精度について評価する。

謝 辞

本研究は JSPS 科研費 JP15K12151, JP16H05932 の助成を受けたものです。ここに記して謝意を表します。

文 献

- [1] Francesco Ricci. Travel Recommender Systems. *IEEE Intelligent Systems*, Vol. 4, No. December, pp. 55–57, 2002.
- [2] Joan Borràs, Antonio Moreno, and Aida Valls. Intelligent tourism recommender systems: A survey. *Expert Systems with Applications*, Vol. 41, pp. 7370–7389, jun 2014.
- [3] Aristides Gionis, Theodoros Lappas, Konstantinos Pelechris, and Evimaria Terzi. Customized Tour Recommendations in Urban Areas Categories and Subject Descriptors. In *[WSDM2014] Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, pp. 313–322, 2014.
- [4] Robert Gyorodi, Cornelia Gyorodi, and Mihai Deridan. An extended recommendation system using data mining implemented for smart phones. *International Journal of Computers and Technology*, Vol. 11, No. 3, pp. 2360–2372, 2013.
- [5] Carlos Lamsfus, David Martin, Zigor Salvador, and Alex Usandizaga. Human-Centric Ontology-Based Context Modelling In Tourism. In *In Proceedings of the international conference on knowledge engineering and ontology development*, pp. 424–434, 2009.
- [6] 大田樹, 奥健太. ラッセルの円環モデルに基づく 観光スポットの感情コンテキスト特徴化大田 樹 奥 健太. ARG 第 11 回 Web インテリジェンスとインタラクション研究会, No. 11, pp. 97–100, 2017.
- [7] 利田柊樹, 奥健太. 投稿写真からの時間コンテキスト特徴化に基づく観光スポット推薦システム. In *DEIM Forum 2018*, pp. D3–3, 2018.
- [8] 篠原修. 景観用語事典. 彰国社, 2007.