

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

Gou KUBOTA

情報メディア学科 2018 年度卒業

1. はじめに

2017 年の訪日外客数が 2,800 万人に達する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。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。

このような背景から、本研究では、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を目的とする。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、景観ラベルの推定を行う。

2. システム概要

2.1 資源への景観ラベルの付与

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 $ri \in R$ は、その資源の名称 $ri.name$ 、地理座標系の位置座標として経度 $ri.x$ および緯度 $ri.y$ をもつ。

資源 $ri \in R$ に対し、景観ラベル si 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 に対し、資源の名称 $ri.name$ 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 ri 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 $ri \in R$ に景観ラベル si を付与する：

(1) クラウドソーシングのワークに、資源 ri 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 1 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、「○○景観」という呼称で回答する。

なお、資源 1 件あたりのワーク数は 5 名とする。

資源 ri 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 ri の景観ラベル si とする。

2.2 資源の地理的特徴化

本研究では、スポットの地理的特徴化を土地利用分類データと標高データに基づき景観の地理的特徴化を行う。地理的特徴化した資源 ri を特徴ベクトル $gi = (10100, 10200, \dots, 11600, emax, emin, eavg)$ で表現する。特徴ベクトル gi 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 1

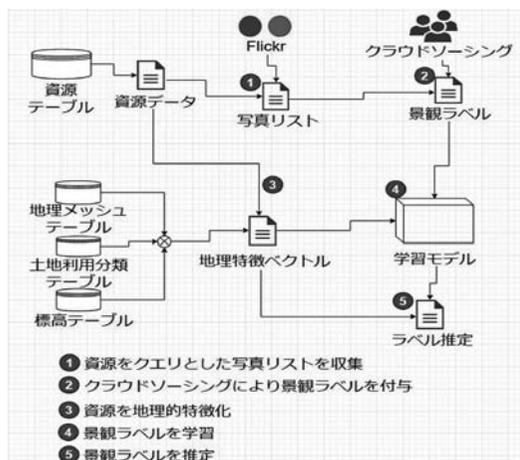


図 1 提案手法の構成図

0100, 10200, . . . , 11600 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_{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。

また, クラウドソーシングにより, 観光スポットに対し景観ラベルを付与する。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。

2.3 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。

2.4 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 s^*_k を推定する。ここで, 資源 r_k の地理的特徴化は 2.2 節と同様である。

3. 評価実験

3.1 データセット

観光スポットの特徴である資源ごとの地理的特徴と, 教師ラベルであるクラウドソーシングによって得た景観ラベルを用いて, 景観ラベルの学習を行う。景観ラベルの学習には SVM による機械学習を行う。機械学習に関しては Python で利用可能なラ

イブラリ的一种である scikit-learn を用いる。学習ではトレーニング用とテスト用のデータを 8:2 の割合で分け, トレーニング用データで学習をし, テスト用データでラベル推薦を行う。

3.2 ラベル推定

17種類ある景観ラベルのラベル推定を行った。結果を見ると, 一番データ数の多い景観ラベルにすべてのデータが集まりよい結果が得られなかった。このことから景観ラベル推定の精度を上げるため, データ数を均等にし, ラベル推定することを今後の課題とする。

4. 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。

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した結果, データ数の一番多い景観ラベルにデータが集まり良い結果は得られなかった。

今後の課題としてデータ数を均等にし, ラベル推定精度の向上を課題とする。