

Flickr を用いた観光スポット自動検出と説明文生成の性能向上を目指して Towards performance improvement for sightseeing spot detection and automatic generation of explanations using Flickr

深津 玲穂
Reiho FUKATSU

広島市立大学 言語音声メディア工学研究室
Speech and Language Research Laboratory, Hiroshima City University

概要 世界中の人々が投稿した写真が大量に蓄積されている Flickr に含まれる緯度経度情報の付いた旅行写真を用いて、観光スポットを自動的に検出し、その説明文を自動生成することを旨とした大加田等の研究[1]の性能向上のための基礎的調査を実施した。本稿では、説明文自動生成における stop-word 導入の効果と、誤訳の影響に関する考察を報告する。

1 はじめに

高性能なデジタルカメラを備えたスマートフォンが幅広く使われるようになったこともあり、日常生活や旅先で気軽に写真を撮影し、様々な Web サービスを活用することで、写真のやり取りを楽しむことができるようになった。

写真を楽しむことができる Web サービスの一つに、写真共有サイト Flickr がある。Flickr には、世界中の人が投稿した膨大な数の写真が蓄積されている。Flickr の写真には、撮影場所の緯度経度情報、撮影日時などの情報が付加されている。

写真は、撮影者の興味や関心を示すものであり、「見る」「食べる」など観光に関わる写真も多く含まれていると期待できる。そこで、大加田等は、Flickr に含まれる緯度経度情報の付いた旅行写真を用いて、観光スポットを自動的に検出し、その説明文を自動生成する研究をしている[1]。

本稿では、大加田等の研究の性能向上を目指して実施した基礎的調査について述べる。説明文自動生成における stop-word 導入の効果と、誤訳の影響に関する考察を報告する。

2 関連研究

Flickr に含まれる写真が撮影者の興味や関心を示すものあることに着目し、Flickr を活用した観光行動の分析[2]や観光情報ツールの研究[3]が行われている。

北村等は、約 17 年間の Flickr データを収集し、撮影者の居住国を推定して、訪日外国人の観光特性を分析している[2]。陳等は、観光施設領域外で観光施設を撮影できる場所を穴場撮影スポットと定義し、Flickr データから見つける手法を提案している[3]。

画像のみを入力として、その内容を説明する自然言語を出力する画像キャプション生成を始めとする画像に関連した言語生成に関する数多くの取組みがある[4]。

日本語を対象とした画像キャプション生成のために開発された大規模なデータセットとその性能評価に関する報告がある[5]、

王らの研究[6]では、ジオタグ（撮影時の GPS 情報）付きの写真を用いて風景カテゴリ識別し、マップに反映する提案が行われている。

3 大加田等の提案手法

観光スポット検出と説明文自動生成の概要は次のとおりである。また、提案手法概略を図 1 にまとめる。

・観光スポット検出

密度に基づくクラスタリング DBSCAN [7]で得られたクラスタを観光スポットと名付ける。

・画像の自動分類

観光に役立つ情報を得るために画像を「見る」「食べる」「買う」「体験する」「泊まる」の 5 つに分類する。

・説明文自動生成

画像の分類結果を使って、観光スポットと画像の説明文を自動生成する。今回はテンプレートに基づく文生成を試みる。



図 1 大加田等の提案手法

3.1 観光スポット検出

密度に基づくクラスタリング DBSCAN により得られたクラスタを観光スポットと名付ける。国内外から多くの観光サイクリストが集まるしまなみ海道を対象に選び、2013 年

から 2019 年までの Flickr データ 15, 407 件を収集した。DBSCAN のパラメータを表 1 に示す。

表 1 : DBSCAN パラメータ

eps	0. 000002
min_samples	3
クラスタ数	951
outlier 数	5503

3.2 分類

藤井等の研究[8]を参考に、「見る」「食べる」「買う」「体験する」「泊まる」に分類する。

Flickr に含まれる緯度経度情報の付いた写真を Google Cloud Vision API にて分析し、ラベルの付与を行う。画像にラベルを付与した例を、図 2. に示す。ラベルごとにその画像の特徴を表す信頼度の score が記されている。



図 2 画像ラベル付与例

Google Cloud Vision API のラベル情報を特徴量として、SVM で分類器を作成した。分類器の性能. を表 2 に示す。

表 2 分類器の評価スコア

	写真数(枚)	精度	再現率	F 値
体験する	101	1. 00	0. 86	0. 93
食べる	101	0. 96	0. 96	0. 96
見る	112	0. 90	0. 74	0. 81
買う	30	0. 00	0. 00	0. 00
泊まる	28	0. 67	0. 50	0. 57

4 説明文の自動生成

大加田等は、観光スポットの説明文と、画像の説明文の自動生成を試みている。ここでは、画像の説明文について考察する。

分類結果ごとにテンプレートを作成し、画像ラベルと組み合わせることで画像の説明文自動生成を行う。テンプレートは、以下のように設定した。「～」には、画像ラベルを一つないし複数個選んで加える。

体験する、食べる：～を楽しむことができる。
 見る：～の景観(観光)を楽しむことができる。
 買う：～で買い物を楽しめる。
 泊まる：～でゆっくりくつろぐことができる。

これらのテンプレートは、分類ごとの特徴を表現するとともに、画像ラベルの出現傾向を考慮し、画像ラベルが挿入された際の言い回しが不自然になりにくいように作成した。

画像の説明文で利用するラベルの選定のために 2 つの手法を提案する。1 つ目は、Google Cloud Vision API から出力される信頼度の score を利用するものである。抽象度の高い語句が高い信頼度になる傾向があるが、画像内の特徴を正確に伝えられることが予想できる。2 つ目は、単語の重要度を表す場面で多く使われる TF- IDF の値を利用することである。幅広い場面で利用可能な語句より特徴的な単語を選択することができるため、前者の手法より具体的な表現が期待できる。

性能向上の検討

Google Cloud Vision API から出力された語句は抽象度が高い語句、観光対象として不適切な語句が含まれる。そのような語句 (food, asphalt など) をあらかじめ Stop-word として排除し、性能の向上を図った。



図 3 Stop-word の例

Stop-word 導入の効果

最も効果が表れたのは「食べる」の画像で Stop-word で抽象度の高い食べ物、料理、飲み物を意味する語句を排除することで okonomiyaki やラーメン、ジャンクフードのように料理名や料理ジャンルが判別できるようになった。「体験する」ではサイクリング、自転車が多く偏りはあるが意図している語句に翻訳することができた。「見る」では自然風景、橋、建築物、飛行機などの乗り物など幅広いジャン

ルの画像が含まれ適切な Stop-word の選定が難しく、不適切な語句が多く残ってしまった。「泊まる」では建物や家具と翻訳されることが多く、適切な文章を生成することができなかった。

「食べる」以外では観光対象として不適切な語句が多く含まれており、適切に語句の排除を行うことができなかった。また、サイクリングや自転車といったしまなみ海道での特性が表れており、地域の特性に合わせて Stop-word の選定を行わなければならないことがわかった。

誤訳の影響

語句ごとに翻訳を行うと train が訓練、track が追跡のように複数の意味を持つ語句が誤訳されてしまう例があった。station のように本来翻訳したい語句に近い意味を持つ語句を一緒に翻訳して意図した翻訳が行えるよう Stop-word で語句を排除した後の語群で翻訳を行った。結果を表 3, 表 4, に示す。

まとめて翻訳を行うことで例に挙げた train, track は実際に翻訳したい語句に翻訳することができた。語群で翻訳したときに意図しない翻訳を行ってしまう事例があった。

表 3 翻訳の結果 1

翻訳前	語句ごと	語群
['Track', 'Train', 'Mode_of_transport', 'Railway', 'Rolling_stock', 'Motor_vehicle', 'Tgv', 'Electric_locomotive']	['追跡', '訓練', '輸送モード', '鉄道', '車両', '自動車', 'TGV', '電気機関車']	トラック、列車、輸送モード、鉄道、ローリングストック、自動車、TGV、電気機関車
['Blossom', 'Flower', 'Spring', 'Cherry_blossom', 'Plant', 'Woody_plant', 'Overhead_power_line', 'Architecture']	['花', '花', '春', '桜の花', '工場', '木質植物', 'オーバーヘッド電力線', '建築']	花、花、春、桜、植物、木質植物、頭上の電力線、建築
['Water_transportation', 'Ferry', 'Boat', 'Luxury_yacht', 'Sea', 'Yacht', 'Ship']	['水輸送', 'フェリー', 'ボート', '高級ヨット', '海', 'ヨット', '輸送する']	水輸送、フェリー、ボート、豪華なヨット、海、ヨット、船
['Noodle', 'Udon', 'Soba', 'Shirataki_noodles', 'Chinese_noodles']	['ヌードル', 'うどん', 'そば', '清い麺', '中国の麺']	ヌードル、ウドン、ソバ、シラタキヌードル、中華麺

表 4 翻訳の結果 2

翻訳前	語句ごと	語群
['Vegetation', 'Daytime', 'Green', 'Cumulus', 'Natural_landscape', 'Hill_station']	['植生', '昼間', '緑', '雲積', '自然の風景', '丘の駅']	植生、昼間、緑、核兵器、自然の風景、丘の駅
['Roof', 'Historic_site', 'Chinese_architecture', 'Architecture', 'Mountain', 'Stock_photography', 'Landscape']	['屋根', '史跡', '中国の建築', '建築', '山', 'ストックフォト', '風景']	屋根、歴史的なサイト、中国の建築、建築、山、ストックフォト、風景
['Torii', 'Temple', 'Place_of_worship', 'Shrine', 'Japanese_architecture', 'Building', 'Shinto_shrine', 'Architecture', 'Leisure', 'Pergola']	['鳥居', '寺', '礼拝の場所', '神社', '日本の建築', '建物', '神社', '建築', 'レジャー', 'パーゴラ']	トリー、寺院、礼拝所、神社、日本建築、建物、神道神社、建築、レジャー、パーゴラ

しまなみ海道と広島広域都市圏（広島、宮島、呉、岩国など）の違いについて

しまなみ海道と広島広域都市圏の共通点として、「見る」に分類される画像が多く、「泊まる」と「買う」に分類される画像が少ない傾向にある。しまなみ海道では「見る」に分類される画像の多くが海、川、橋と自然風景に分類されるものであった。広島広域都市圏では原爆ドーム、厳島神社、錦帯橋等の観光地の画像や路面電車、米軍基地等の地域の特性を表す画像が多く、人工物の画像が多かった。

観光地とされている場所は「見る」から「原爆ドーム」、「厳島神社」など小分類を検討する。

まとめ

Flickr に含まれる緯度経度情報の付いた旅行写真を用いて、観光スポットを自動的に検出し、その説明文を自動生成することを目指した研究の性能向上を検討した。この検討結果を参考にして、広島、宮島、呉、岩国など広島広域都市圏を対象に、高性能で使い勝手の良い情報提供と説明文生成が可能となるように研究を進める予定である。

5 参考文献

- [1] 大加田華実, 石野亜耶, 竹澤寿幸, Flickr を用いた観光スポット検出と説明文自動生成, 電子情報通信学会言語理解とコミュニケーション研究会, 2022年3月.
- [2] 北村武士, 本間健太郎, 今井公太郎, Flickr のジオタグ付き写真データからみる日本全土の観光特性 居住国推定とタグクラスタリングによる訪日外国人の興味分析, 日本建築学会計測系論文集, 第 84 巻, 第 755 号, pp. 187-197, 2019年1月.
- [3] 陳嘉穎, 新妻弘崇, 太田学, Flickr を利用した穴場撮影スポットの発見の一手法, DEIM Forum 2019, F1-6, 2019.
- [4] 牛久祥孝, 画像に関連した言語生成の取組み, 人工知能学会誌「人工知能」, Vol. 34, No. 4, pp. 483-491, 2019.
- [5] 吉川友也, 重藤優太郎, 竹内彰一, STAIR Captions: 大規模日本語画像キャプションデータセット, 言語処理学会 第 23 回年次大会 発表論文集, pp. 537-540, 2017.
- [6] 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 3588-3592 (Dec. 2011).
- [7] Martin Ester, Hans-Peter Kriegel, Jorg Sander, Xiaowei Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases in Noise, " Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), pp. 226-231, 1996.
- [8] 藤井一輝, 難波英嗣, 竹澤寿幸, 石野亜耶, 奥村学, 倉田陽平, 旅行者の行動分析のための旅行プロジェントリの属性推定, 観光情報学会誌「観光と情報」, 第 13 巻, 第 1 号, pp. 83-96, 2017.