

Deep Learning を用いた自然系観光地に対する印象評価推定 AI の作成と検証

- 観光口コミサイトの画像とコメントを対象に -

大嶋 康平

1. はじめに

1. 1. 研究の背景と目的

これまで、訪日外国人旅行者数やその消費額は近年急速に増加してきた。それにより、日本経済における観光の存在感も高まっていた。しかしながら、2020年（令和2年）1月以降、新型コロナウイルス感染症の世界的な拡大により、観光需要は大きく減少している。今後、コロナ以前の外国人旅行者数に戻るまで、長い時間がかかることが予想される。したがって、日本人に特化した対応をすることがまず第一歩である。

しかし、今現在、旅行者にアンケートを取るのには、困難な状況である。そこで近年様々な人工知能技術を用いた画像認識システムの実用化が進んでおり、とくに Deep Learning（深層学習）は、国際的な画像認識コンテスト「ILSVRC」^(注1)において従来の画像認識手法を大幅に上回る認識精度を実現し、現在の人工知能ブームの火付け役となった。また、画像認識の技術は数多くの分野に応用されており、情報技術の普及に伴って、都市建築分野においても、これまで以上に画像を取り扱う機会が増えた。

そこで、本研究は都市計画分野における画像認識技術を用いた先駆的な研究として、「日本の観光地の画像」と「そのコメント」を学習させ、自然系観光地の評価を推定する AI の適用可能性を明らかにすることが目的である。

1. 2. 既往研究

関連する既往研究として、山田ら⁽¹⁾は、異なる都市の街路の街並み画像と訪問意欲の分類・推定に Deep Learning を適用している。また、VR 技術を併用した研究も行われており、瀧澤ら⁽²⁾は VR 空間で実施した印象評価実験を VR 空間の 360 天球画像から推定することに Deep Learning を適用している。しかし、今現在、コロナウイルスの感染拡大を考慮して接触を伴う印象評価実験を行わずに、観光地口コミサイトからスクレイピングした日本の自然系観光地の画像に絞って評価の分類・推定に Deep Learning の適用を試みた研究は筆者が知る限りでは見当たらず、この点が本研究の新規性に該当する。

1. 3. 研究方法

本研究では、既往研究を参考に画像が約 1000 件となるように、また、観光地によってスクレイピングされるコメントの件数が偏らないように観光口コミサイト TripAdvisor^(注2) から日本人の口コミが多く投稿された 6 地域の上位 10 位までの観光地のコメントを 100 件ずつスクレイピングする。その中で画像が含まれていないものを取り除き、画像について構成要素分析を行う。その結果、自然系の構成要素が 50% 以上含まれていると認められたものについて、自然系観光地の画像として、そのコメントを「Positive」と「Negative」の 2 クラスに分類する。さらに、コメントを感情分析した 3 つのクラスに分類する。最後に、その結果を Deep Learning を用いて、AI に学習させ、その精度を検証する。

表 1 対象観光地

東京都		北海道		京都府	
1	JAL 工場見学	1	神威岬	1	三十三間堂
2	明治神宮	2	旭岳	2	禅院寺永観堂
3	千鳥ヶ淵	3	ニッカウキスキー 余市蒸留所	3	青蓮院
4	根津美術館	4	函館山	4	伏見稲荷大社
5	新宿御苑	5	礼文島	5	東福寺
6	東京国立博物館	6	旭山動物園	6	仁和寺
7	江戸東京たてもの園	7	上野ファーム	7	東寺
8	東京タワー	8	大倉山ジャンプ 競技場	8	鞍馬山
9	吉祥寺	9	滝野スズラン 丘陵公園	9	大河内山荘庭園
10	羽田空港第 2 ターミ ナル展望デッキ	10	知床五湖	10	桂離宮
大阪府		沖縄県		神奈川県	
1	サントリー山崎 蒸留所	1	ニシ浜ビーチ	1	YOKOSUKA 軍港 巡り
2	ザ・シンフォニー ホール	2	沖縄美ら海水 族館	2	ポーラ美術館
3	万博記念公園	3	与那覇前浜 ビーチ	3	大さん橋
4	なんばグランド花月	4	八重干瀬	4	報国寺 (竹寺)
5	インスタントラーメン 発明記念館	5	座間味島	5	長谷寺
6	造幣博物館	6	伊良部大橋	6	横浜みなとみらい 21
7	住吉大社	7	西表島	7	キリンビール 横浜工場
8	キッズプラザ大阪	8	コンドイビーチ	8	円覚寺
9	法善寺	9	果報バンタ	9	三溪園
10	大阪市中央公会堂	10	はての浜	10	明月院 (あじさい寺)

2. 研究概要

世界的観光口コミサイト Tripadvisor において、日本人の口コミ数が多かった東京都、北海道、沖縄県、神奈川県、大阪府、京都府の評価点数や投稿件数から出されたランキングの上位 10 地域の観光地に対して投稿された口コミの新しいものから 100 件ずつスクレイピングした合計 6000 件のコメントを用いる。スクレイピングには、Google の拡張機能である Web Scraper を用いた。Web Scraper は Web サイト内の情報を取得できるスクレイピングツールである。

次に、構成要素の分析を行った。スクレイピングによって得られた 3633 枚の画像の中には、食べ物や建物の写真など今回の研究に適さない画像も含まれる。そこで、まず画像の構成要素を分類する。今回利用した画像構成要素分析アルゴリズムは、MXNet^(注3) で事前に学習させたモデルを用いて構築した。このアルゴリズムは、画像に含まれる景観要素とその割合を判断し、その上で、画像に表現された景観の種類（自然景観か文化景観か）を判別するものである。

次に、コメントの感情分析を行った。まず、コメントを「Positive」と「Negative」の 2 クラスに分類する。次に、さらに詳しく「Positive」を「happy」と「surprise」の 2 つに、「Negative」を「sad」に、合わせて 3 クラスに分類した。今回利用したのは、日本語の Sentiment Analysis^(注4) である sentiment_ja を用いた。感情分析とは、文全体が持つ感情の特徴（専門的には、感情極性と呼んだりします）を予測するタスクであり、文書分類器を用いて解かれるものである。

最後に、Deep Learning を用いて、上述の結果を AI に学習させ、評価の推定を行った。具体的には、予測モデルの構築に Apple Core ML architecture^(注5) を用い、画像とともに取り出したコメントを感情分析した結果をもとに、自然系観光地の画像をラベル付けし、その評価を推定する AI の作成を行った。

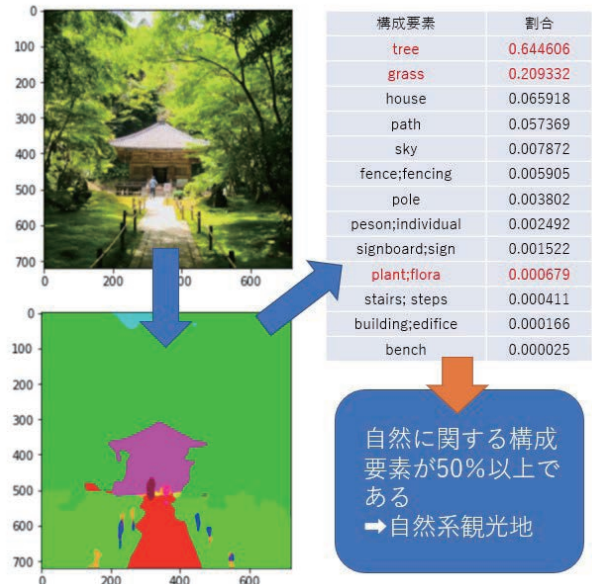


図1 構成要素分析アルゴリズム



図2 テキスト感情分析アルゴリズム



図3 スクレイピング画像例 (左 6 枚 Positive 右 6 枚 Negative)

3. 学習結果（2クラス分類）と評価

学習データセットは、上述した処理を経て、6000件のコメントから3633枚の画像について構成要素分析を行い、自然系の観光地と判断された1008枚とした。本章のクラスは「Positive」「Negative」という評価の2クラスである。学習の推移と結果を示す。

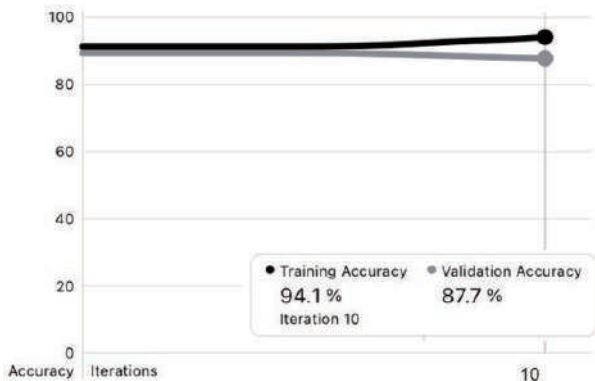


図4 2クラス分類AIの学習結果

グラフの縦軸は正確性であり全画像枚数のうち何枚を正しいクラスに推定したかを示している。一方は作成する際に使用した学習用のデータセットに対する正確性であり、他方はAIがランダムに抽出したデータセットに対する正確性を示している。横軸は学習の回数である。既往研究を参考に学習回数は10回とした。具体的な結果を示す。

表2 2クラス分類AIの結果評価

		Item Count	Precision	Recall
Training Data	Positive	815	96%	97%
	Negative	78	68%	58%
Validation Data	Positive	58	90%	93%
	Negative	7	20%	14%
		AI判定		合計
		P	N	
実際クラス	P	844	29	873
	N	29	56	85
Accuracy(正確性)		0.940595	F-Measures	0.967113
Precision(予測適合率)		0.974233	Kappa係数	0.6256047
Recall(再現率)		0.960097		

Accuracyは正確性であり、全画像枚数のうち何枚を正しいクラスに推定したかを示す。Precisionは予測適合率であり、あるクラスにおいて推定した枚数に対して何枚を正しいクラスに推定したかを示す。Recallは再現率であり、あるクラスにおいて、実際の枚数に対して何枚を正しいクラスに推定したかを示す。各々の値を見ると、Accuracyが94%、Precisionが97%、Recallが96%であり、比較的高い精度であるといえる。なおRecallとPrecisionはトレードオフ

の関係にあることから、このような推定精度の検証には両者の調和平均であるF-Measuresも用いられる。F-Measuresの値は、約96%であり、高い精度が得られた。次に統計的な精度評価を示す。用いた統計指標はKappa係数である。Kappa係数は医療統計や心理統計に用いられている統計量であり、値は-1から1までの値を取り、数値が1に近いほど精度が高いといえる。今回のKappa係数は0.6256となり、医療統計の分野の知見を引用すれば、実質的に一致しているとみなされる値である。計算式を以下に示す。

$$Accuracy = (TP + TN) / (TP + FP + FN + TN)$$

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

TP：真陽性 FP：偽陽性 FN：偽陰性 TN：真陰性

$$F - Measures = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

$$Kappa \text{ 係数} = \frac{P_o - P_c}{1 - P_c}$$

P_o：一致率
P_c：偶然の一致率

以上のように本章では、AIの建築・都市に関わる画像への萌芽性について、観光口コミサイトのコメントから自然系観光地の評価を推定するAIについては、F-Measuresの観点からは96%、Kappa係数は0.6256という精度であることを明示した。学習モデルは日々進化しており、より正確な推定をするAIを作成する余地はあると思われるが、統計的にも比較的高いといえる精度を確認したことから、本研究は観光口コミサイトのコメントから自然系観光地の評価を推定するAIを作成することができたといえる。

4. 学習結果（3クラス分類）と評価

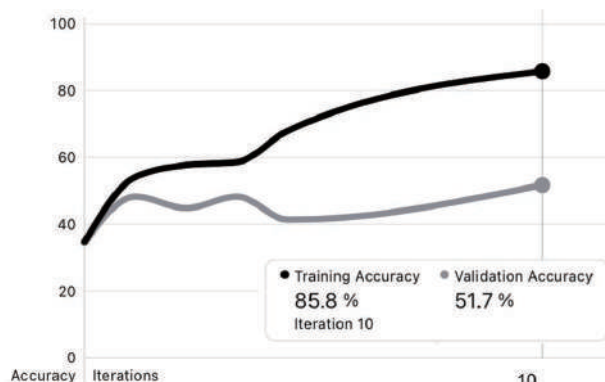


図5 3クラス分類AIの学習結果

本章では前章にて作成した推定AIを援用することで、より細かな感情評価の推定が可能かを検証する。本章のクラスは「happy」「surprise」「sad」という感情評価の3クラスである。画像は前章のものに加えて、対象観光地の評価点数の低いものについて追加でスク

レイピングと構成要素分析を行い、各クラスでデータ数に大きく偏りが出ないようにデータ数を調整したものをを用いた。内訳は「happy」が120枚、「surprise」が118枚、「sad」が122枚である。前章と同様に学習回数は10回とする。

図5のグラフを見ると、Training Accuracyは85.8%、Validation Accuracyは51.7%という値をとった。前章のものと比較すると、その値は前者が8.3%、後者が36.0%下がっている。

表3 3クラス分類AIの結果評価

		Item Count	Precision	Recall	
Training Data	happy	111	81%	85%	
	surprise	108	87%	82%	
	sad	112	89%	90%	
Validation Data	happy	9	45%	56%	
	surprise	10	50%	30%	
	sad	10	58%	70%	
happy	accuracy	0.880556	macro平均	accuracy	0.88333333
	recall	0.825		recall	0.82441187
	precision	0.818182		precision	0.82537371
surprise	accuracy	0.875	F-Measures	happy	0.82157676
	recall	0.771186		surprise	0.80176211
	precision	0.834862		sad	0.84920635
sad	accuracy	0.894444	F-measures		0.82418174
	recall	0.877049	Kappa係数		0.6222846
	precision	0.823077			

3つのクラスの分類タスクを評価する際には、クラスを一つ選びPositiveとし、それ以外をNegativeとして評価値を計算した後に、それらの平均値をとることで全体の評価値を計算する。このときの平均のとり方としてmicro平均とmacro平均がある。今回、データ数に偏りは無いため、macro平均を用いた。表3の具体的な結果を見ると、Accuracyが88%、Precisionが83%、Recallが82%であり、前章のものに比べるとその精度は劣るものの比較的高い精度であるといえる。また、RecallとPrecisionの調和平均であるF-Measuresの値は82%となり、比較的高い精度が得られたといえる。さらに、統計的な精度評価であるKappa係数は、0.6223であり、医療統計の分野の知見を引用すれば、実質的に一致しているとみなされる値である。

以上のように本章では、前章の成果に加えて、観光口コミサイトのコメントから自然系観光地の感情評価を推定するAIについては、F-Measuresの観点からは82%、Kappa係数は0.6223という精度であることを明示した。学習モデルは日々進化しており、より正確な推定をするAIを作成する余地はあると思われるが、統計的にも比較的高いといえる精度を確認したことか

ら、本研究は観光サイトのコメントから自然系観光地の感情評価を推定するAIを作成することができたといえる。

5. 結論

5.1. まとめ

本研究では、まず、観光口コミサイトから画像とコメントをリンク付けしてスクレイピングし、構成要素の分析を行った。次に、Sentiment Analyzerを用いて分類したコメントと画像をDeep Learningを用いて作成したAIに学習させ、その評価推定AIの適用可能性について検証した。

まず「Positive」「Negative」の2クラスの推定において比較的高い精度の推定AIを作成することができた。その精度を統計的に表すと、F-Measuresが約96%、Kappa係数が0.6256である。

次に「happy」「surprise」「sad」の3クラスの感情評価の推定においては、2クラス分類に比べて精度は劣るものの比較的高い精度の推定AIを作成することができた。その精度を統計的に表すと、F-Measuresが約82%、Kappa係数が0.6223である。

5.2. 今後の課題

学習モデルは日々進化しており、物体認識においてはさらに高い精度の結果も報告されている。印象の推定に関する事例は多くないが、適用が進むと考えられる。本研究においてもさらに高い精度をもつ推定AIを作成することを今後の課題とする。また、今回の研究を踏まえ、対象の観光地を変えた上での検証、感情分析を細かく行った複数クラスでの検証、異なる観光サイトでの検証、SNS等のコメントでの検証、アウトバウンドの評価推定AIの作成検証に取り組みたい。

【補注】

- 注1) ImageNet Large Scale Visual Recognition Challengeの略
- 注2) コンピューターによる画像認識技術に関するコンペティション形式の研究集会
- 注3) 世界最大級の旅行プラットフォームで、利用者の口コミや評価などが閲覧できる
- 注4) フル機能のディープラーニングフレームワークであり、「畳み込みニューラルネットワーク(CNN)」「長短期メモリネットワーク(LSTM)」などをサポートしている
- 注5) 一般的にテキストに含まれる感情を予測する多値分類の文書分類タスクを指す
- 注6) 独自のカスタムレイヤを定義して使うことのできるモデル

【参考文献】

- 1) 山田悟志、大野耕太郎: Deep Learningを用いた印象評価推定AIの作成と検証 - 街並み画像の街路名と訪問意欲を対象に - 日本建築学会計画系論文集 第84巻 第759号 1323-1331 2019
- 2) Atsushi Takizawa and Airi Furuta: 3D Spatial Analysis Method with First-Person Viewpoint by Deep Convolutional Neural Network with Omnidirectional RGB and Depth Images, eCAADe 2017, Sapienza University of Rome, Rome, Italy, pp. 693-702, 2017. 10
- 3) 山田悟志、大野耕太郎: Deep Learningを用いたデザインAIの作成と検証 - 街並みと建築物外観の画像作成を対象に - 日本建築学会計画系論文集 第85巻 第770号 987-995 2020
- 4) 斎藤康毅: ゼロから作るDeep Learning - Pythonで学ぶディープラーニングの理論と実装 オライリージャパン 2016
- 5) Yujie Ren, Shichen ZHAO: A METHOD ON RECOGNIZING THE COGNITIVE PREFERENCES OF URBAN LANDSCAPE BASED ON MACHINE LEARNING
- 6) デービッド・アトキンソン: 世界一訪れたい日本の作り方 東洋経済新報社 2017
- 7) Klaus Krippendorff: Content Analysis: An Introduction to Its Methodology (2nd Edition), Sage Publications, 2003