

# SNS 投稿画像を用いた環境問題に潜在的に関心をもつ ユーザー抽出に向けた初期検討

柏木 直人<sup>†</sup> 鈴木 積規<sup>†</sup> 李 貞憲<sup>‡</sup> 池田 大輔<sup>\*</sup>

九州大学大学院システム情報科学府<sup>†</sup> 九州大学大学院理学研究院<sup>‡</sup>

九州大学大学院システム情報科学研究院<sup>\*</sup>

## 概要

オーバーツーリズムは、ゴミのポイ捨て等の環境問題を引き起こしているが、未だ決定的な打開策は確立されていない。もし、潜在的に問題意識を持つ人に働きかけることができれば、効率良く啓発が行えると考えられる。その第一歩として、本研究では、SNS 画像に着目し、「何に焦点を当てて撮影しているか」に基づいて、ユーザーの環境に対する潜在的な問題意識を判別する課題を考える。従来の多くの研究は、分類基準が明確な画像を用いて、分類精度を実験的に評価している。一方、本研究では、撮影者の意識を考慮するため、分類基準が曖昧な画像を用いて自動判別の評価・検討を行うことを目的とする。そのため、環境に対する問題意識を判別できるような、明確なラベルの定義を考案し、それを基にアノテーションを行う。合計 154 枚の画像を用いて、ホールドアウト法により分類精度を評価した結果、65.38%を示した。

## 1. はじめに

2003 年の観光立国宣言を機に、わが国はインバウンド(訪日外国人客)振興策に取り組んでおり、東京オリンピック・パラリンピックが開催される2020年次の目標「インバウンド集客4,000万人」の実現も視野に入りつつある。一方で、国内外の観光客が一部に集中するオーバーツーリズムと呼ばれる弊害も明らかになってきている。オーバーツーリズムは様々な弊害を引き起こしており、わが国に先行して海外でも観光公害が顕在化している[1]。ゴミのポイ捨て等の環境問題もその一つであり、様々な対応策が取られているが、未だ解決には至っていない。

もし、潜在的に問題意識を持つ人に働きかけることができれば、効率良く啓発が行えると考えられる。その最初の取り組みとして、SNS 画像

に着目し、「何に焦点を当てて撮影しているか」に基づいて、ユーザーの環境に対する潜在的な問題意識を判別することを試みる。従来研究の多くは、分類基準が明確な画像を用いるが、本研究では、撮影した写真に反映されるユーザーの環境に対する問題意識を考慮するため、分類基準が曖昧であり、挑戦的な課題と言える。環境に対する問題意識を判別できるような、明確なラベルの定義を考案し、それを基にアノテーションを行った。154 枚の画像を入力データとして用い、分類精度を確認するための検証方法としてホールドアウト法で精度を評価した結果、65.38%を示した。

## 2. 問題設定

本研究では、まず、観光公害が顕在化している都市に着目し、ユーザーが「何に焦点を当てて撮影しているか」に基づいて、環境に対する潜在的な問題意識を判別する課題を考える。図 1 に示すように、ゴミや自然、植物に焦点を当てて SNS に投稿しているユーザーは環境に対して問題意識を持っていると考えられる。



図 1: 環境問題に意識のあるユーザーの写真と(左図)環境に対し潜在的に問題意識を持ちそうなユーザーの写真(右図)

画像データを作成する際、ユーザーの環境に対する潜在的な問題意識が反映されるようなラベルの定義を考え、それを基にアノテーション

Initial study for extracting users with potential interest in environmental problems using SNS post images

<sup>†</sup> Naoto Kashiwagi Tokinori Suzuki

<sup>†</sup> Graduate School of Information Science and Electrical Engineering, Kyushu University

<sup>‡</sup> Jounghun Lee

<sup>‡</sup> Faculty of Science, Kyushu University

<sup>\*</sup> Daisuke Ikeda

<sup>\*</sup> Faculty of Information Science and Electrical Engineering, Kyushu University

を行う。ラベルの定義は次のように設定する。ラベル 1: 自然や観光名所, 植物に焦点が当たっている。ラベル 2: ラベル 1 とラベル 3 ではないがその中間に位置する。ラベル 3: セルフィーなど人にしか焦点が当たっていない。ラベル 4: その他(食べ物など, 観光や自然とは無関係なもの)。これらの定義に従ってラベル付けした画像データを用いて, ユーザーの環境問題への興味や関心を考慮した自動判別の評価・検討を行う。

### 3. 実験

アノテーションした画像データを入力として, 事前学習済みネットワーク AlexNet[2]で画像分類を行う。以降で, データセット, 評価実験の結果と考察について述べる。

#### 3.1 データセット

フリー百科事典ウィキペディア(Wikipedia)で, オーバーツーリズムが問題になっている地域<sup>\*1</sup>を実験の対象とする。富士山, 白川郷, イースター島など観光公害が顕著化している都市があるなか, 本研究では, マチュピチュを取り上げデータセットを作成する。

まず, Instagram からマチュピチュの画像とコメントが含まれるポストを 1029 件収集した。そのうち 300 件に対して, 考案したラベルの定義を基に, 3 人にアノテーションを行ってもらった。画像とコメントに含まれる要素を反映させて, 作業者が最も近いと思うラベルを画像全体のラベルと設定する。また, 入力に用いる画像データは, 3 人によるアノテーションが一致した合計 154 枚のみを使用した。3 作業者の一致率を評価する指標としてカッパ係数[3]を用いた。3 作業者のカッパ係数は 0.356 と低めであった。各ラベルの一致した画像枚数を表 1 に示す。

表 1 データセット詳細

ラベル	1	2	3	4	計
データ数	63	42	17	32	154

#### 3.2 性能評価

画像データを 70%の訓練データと 30%のテストデータに分割し, ホールドアウト法で検証を行った。実験の詳細を表 2 に示す。

表 2 実験詳細

最適化手法	sgdm
入力サイズ	227 × 227
ミニバッチサイズ	10
最大エポック	5

ホールドアウト法で性能評価を行った結果を

表 3 に示す。

表 3 実験結果

訓練データ精度	検証データ精度
96.2 %	65.38 %

#### 3.3 考察

画像データが 154 枚と少ないため, 十分な教師あり学習を行うことができず, 精度が低い。訓練データ精度と検証データ精度に乖離があることから, 過学習であることが考えられる。また, 3 人のアノテーションが一致した画像が少なかったことから, ラベル付けの定義を見直す必要性がある。

#### 4. おわりに

本研究では, アノテーションコストが大きかったため, 十分なデータ枚数を用意できなかった。そのため, データオーグメンテーションや半教師あり学習で画像データの枚数を増やす必要がある。また, アノテーションの一致率が低かったので, ラベル付けの際は, 作業者の多義性を考慮したマルチラベルにすることが有効な手段であると考えられる。本論文では, オーバーツーリズムが顕著化している都市のうち, マチュピチュのみを取り上げて検証を行ったが, 他の観光地についても同様の実験を行う必要がある。

#### 謝辞

本研究のデータ作成・実験に協力してくれた池田研究室一同に感謝します。

#### 参考文献

- [1] 高坂 晶子, 「求められる観光公害(オーバーツーリズム)への対応: 持続可能な観光立国に向けて」, 『JRI レビュー』, vol. 6, no. 67, 97-123, 2019.
- [2] A. Krizhevsky, I. Sutskever, and G. Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”, In *Proc. of NIPS 2012*, pp. 1097-1105, 2012.
- [3] J. L. Fleiss. “The measurement of interrater agreement”, In J. L. Fleiss et al. (eds.) *Statistical methods for rates and proportions*, Wiley, pp. 212-236, 1981.

\*1 <https://ja.wikipedia.org/wiki/観光公害#オーバーツーリズム>