Electronic Preprint for Journal of Information Processing Vol.30

Regular Paper

Grasping Users' Awareness for Environments from their SNS Posts

Tokinori Suzuki[1,a)] Naoto Kashiwagi[2] Jounghun Lee[3] Kun Qian[4] Daisuke Ikeda[1,b)]

Received: May 31, 2021, Accepted: December 3, 2021

Abstract: Overtourism has a negative impact on tourist sites all over the world. Serious problems are environmental issues, such as littering, caused by the rush of too many visitors. It is important to change people's mindset to be more environmentally aware for improving this situation. In particular, if we can find people with a high awareness about environments, we can work effectively to promote eco-friendly behavior by taking them as the start. However, grasping individual awareness is inherently difficult. For this challenge, we utilize SNS data, which are available in large volume, with a hypothesis that people's subconsciousness influences their posts. In this paper, we address two research topics for grasping such awareness. First, we propose a classification task, in which a system is given users' SNS posts about tourist sites, and classifies them into types of their focuses. Experimental results show widely-used classifiers can solve the task at about 0.84 of accuracy using our created dataset. Second, we investigate the relation of the focuses and such awareness with a questionnaire survey targeting over 2,700 people, and show that users' awareness influences focuses of SNS posts with both of a statistical analysis and an analysis using real-world data.

Keywords: overtourism, ecotourism, prediction on eco-friendly users, environmental awareness, social network services

1. Introduction

As international airfares become cheaper, the tourism market becomes bigger, and cheap accommodation options have emerged [13], many visitors are going to a popular tourist site at the same time, resulting in overtourism. Overtourism has had a negative impact on various things at tourist sites. In Jeju, for example, its once-pristine environment has been seriously damaged after it became popular with outsiders, and it is reported that there is more trash and traffic jams [16]. Among various issues caused by overtourism, we focus on environmental issues caused by overtourism in this paper.

Due to its depth of overtourism, there seems to be no quick solution for environmental issues caused by overtourism. Instead, as a slow and steady approach, we need to change our mindset about the issues. However, to change the mindset of an individual is difficult. In particular, people who are highly conscious of nature and environmental issues, we call them *eco-friendly people* who are practicing some environmental protection activities, such as avoiding littering, and spreading their practices for others by social networking sites (SNSs) for example. Such eco-friendly people are in the minority, and it is much more difficult to change the mindsets of individuals of the majority.

Even in the majority of people, there are many levels of awareness about environmental issues. We assume there exists a level of people who are comparatively highly conscious but not so high as to practice some environmental protection activities now, and possibly can be eco-friendly people in the future. We call such people *potentially eco-friendly people*. If we can find potentially eco-friendly people, we can promote eco-friendly behaviors, gradually convincing the majority with potentially eco-friendly people as a clue, shown by a result in sociology [2].

For this approach, it is important to grasp an awareness of people for such issues. However, it is inherently difficult to directly grasp such awareness of many people due to enormous costs. Instead of doing so, our idea is to utilize data on SNSs to discern such awareness inspired by recent studies using a large amount of SNS data for understanding people's awareness [15], [17], [22]. SNSs are commonly used on a daily basis, and the posted pictures and comments may reflect the potential interests and concerns of users. For example, **Fig. 1** shows a post uploaded to Instagram [*1], which is a popular SNS site to share images and videos. The picture of the post depicts a person picking up litter on Jeju Island and its comments says that #zerowaste and #Jeju trash. From this picture, the user of this post seems to be sad to see Jeju Island is being polluted by litter.

Let us look on more general examples. Users who upload pictures like the post of **Fig. 2** a) seem to be interested in nature and environments of tourist sites because they make efforts to take pictures of nature in tourist sites and share them in the SNS site. On the other hand, users who upload pictures like the post of Fig. 2 b) take pictures trying to make themselves good looking, and thus seem to have little interest in nature or environments of

---

[1] Kyushu University, Fukuoka 819–0395, Japan
[2] Deloitte Tohmatsu Consulting LLC, Chiyoda, Tokyo 100–8361, Japan
[3] Seoul National University, Gwanak-gu, Seoul 08826, Korea
[4] Fukuoka University, Fukuoka 814–0180, Japan
[a)] tokinori@inf.kyushu-u.ac.jp
[b)] daisuke@inf.kyushu-u.ac.jp

[*1] https://instagram.com/

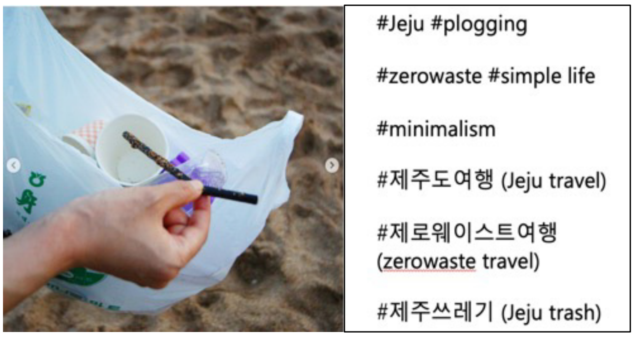


Fig. 1 A post about littering on Jeju island.

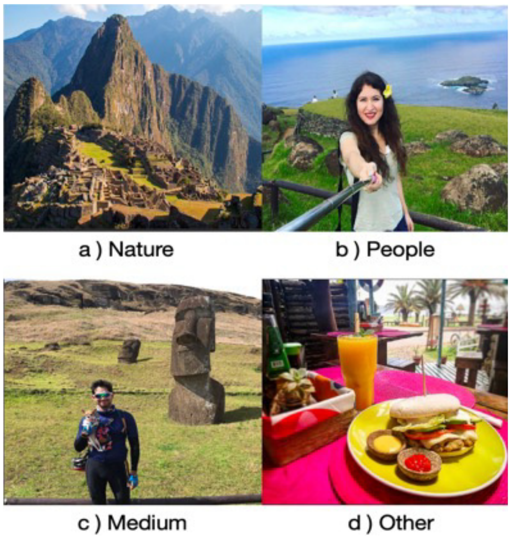


Fig. 2 Our defined labels and example posts assigned with the labels.

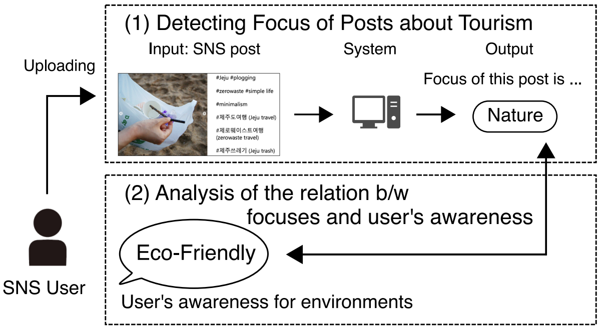

(1) Detecting Focus of Posts about Tourism
Input: SNS post
System
Output
Uploading
Focus of this post is ...
Nature
(2) Analysis of the relation b/w focuses and user's awareness
Eco-Friendly
SNS User
User's awareness for environments

Fig. 3 Two research topics for estimating users' awareness.

tourist sites. Therefore, we set up the hypothesis that focuses of contents in SNS posts, such as images and comments, reflect the users' awareness.

The goal of our research including this paper is to find potentially eco-friendly people from data on SNSs. We think that two main research topics illustrated in **Fig. 3** should be studied for achieving this goal. The first topic is about developing systems detecting focuses of SNS posts shown in Fig. 3 (1). As mentioned above, the detected focuses will be hints for grasping users' interests. However, after such focuses are successfully detected, there is a gap between detected focuses and users' awareness because all the posts of a user do not have to show his/her awareness even if the user is eco-friendly. Eco-friendly users take photos such as Fig. 2 b), and may open such photos on their SNS. Therefore, the second research topic is to investigate the relation between users' focuses of their posts and their awareness for environments in order to fill in the gap, as shown in Fig. 3 (2).

In the previous paper [9], we introduced four focus types of SNS posts, which are used as labels to classify posts about tourist sites: "Nature", "People", "Medium", and "Other". Example posts of each label are shown in Fig. 2. Then, we formalized a research task called Detecting Focuses of Posts about Tourism (DFPT), in which a system is given a post consisting of an image and a comment as an input, to classify the input into one of the four labels. That corresponds to one topic in Fig. 3 (1)). We proposed simple classification methods with supervised learning frameworks, and presented the performances as baselines for the DFPT. However, the previous study is limited to laying the basis for evaluating the DFPT, thus formalizing the task and creating a dataset. Since the previous study lacked user studies about their intentions of posting, the relation between the focuses and users' awareness for environments is unclear.

To establish a correspondence between users' focuses of SNS posts and their awareness for environments, which is not addressed in the previous study, we investigate the relation between the focuses and their awareness through a questionnaire in this paper. That corresponds to the other topic in Fig. 3 (2). We conducted a large-scale questionnaire covering 2,743 people of various ages in order to analyze the relation. With the investigation, we show that there is a statistical significance between levels of users' awareness for environments and selecting focuses of photos when they visit a tourist site. Furthermore, to estimate real users' awareness, proportions of the focuses in all the posts of their history should be different from other users with different levels of awareness, and the developed classifier can work for real-world data. To check these, we conducted an analysis using real-world data of Instagram users to estimate proportions of the focuses in all of their posts using our classification methods. We show investigation results that the proportions are different among each group of different level of the awareness.

The contribution of this paper is the following three folds.

(1) We formalize a new task called Detecting Focuses of Posts about Tourism, and create a new dataset for evaluating the task.

(2) We conduct an experiment of the DFPT with three CNN classifiers for images of Instagram and with two classifiers using the LSTM and the BERT for the comments. The results indicate that the task can be solved at about accuracy of 0.84 with a simple classification method, which are baselines for future researches.

(3) We conduct a large-scale questionnaire investigation targeting 2,743 people to analyze focuses of SNS posts and users' awareness. With this investigation, we confirm that their awareness influences focuses of photos uploaded to SNSs with a statistical test. We also show that it is possible to estimate SNS users' awareness for environments within the task framework of the DFPT with an Instagram real-world data analysis.

2. Related Work

As described, we try to identify users' hidden awareness from their data on SNS in order to support tourism. In this section, we briefly review the literature from the following two viewpoints: identifying users' hidden awareness on their SNS data and studies for supporting tourism utilizing SNS.

First, there exist several studies [1], [15], [17], [18], [22] similar to our study in terms of input and output. That is, given SNS data as an input, the studies try to identify users' hidden awareness on their SNS data. Xi *et al.* classify Facebook images of politicians into political ideologies of the U.S. politicians, such as liberal, conservative and political parties [22]. Ofli *et al.* quantify SNS users' perception for food by measuring the difference of users' uploaded Instagram images and their assigned tags, which is called a perception gap, and analyze the correlation between the gap and the health statistics in the U.S. [17]. Liu *et al.* predict whether harassment comments will be issued on message threads of an Instagram post or not, using the textual comments [15]. This line of studies is similar to ours on using SNS data. However, their focus is only on images or on text of comments. In our DFPT, not only images or text but also both data would contain important information for identifying users' focuses because the posts of Instagram consist of images and users' comments.

Second, since our motivation of this study is enhancing tourism, the existing literature for supporting tourism using SNS shares a similar goal on tourism with ours. Supporting local tourism association has been attempted [18], [20]. Suzuki proposes a clustering method of tourist sites by analyzing users profile information of twitter utilizing following relations of users to local tourism associations [20]. This aims at promoting the adoption by associations of a successful PR strategy with reference to other tourist sites within the same group. Ohkubo and Muromachi analyze the reviews of Tripadvisor [*2], which is the largest review site for things related to traveling such as hotels, places by text mining in order to grasp tourists' image for tourist sites [18]. Then, supporting foreign tourists with SNS data has been studied [14]. Foreign tourists may have difficulties communicating with people when visiting a country. For this problem, the system displays a map with geotagged SNS pictures of recommended sites for foreign tourists in order to assist them in planning their travel. These existing studies focus on supporting tourism. However, to the best of our knowledge, there are no studies attempting to improve the situation caused by overtourism using SNS.

3. Dataset for the DFPT

Since there are no evaluation datasets for the DFPT task, we created one. In this section, we explain the definition for classification labels of multimedia SNS posts about tourist sites based on their focuses. We then describe the annotation process and the dataset developed.

3.1 Classification Labels

First of all, we need classification labels distinguishing levels of people's focuses for nature and environmental issues from their SNS posts. For this purpose, we defined four labels based on their focuses as shown in **Table 1**. For the label definition, we preliminarily investigated several hundreds of posts uploaded to Instagram, such as Kyoto in Japan and Jeju island in Korea. After the investigation, we identified the three labels, Nature, People and Medium, which show the levels of the people's awareness. Nature label is for the posts from which a person might be interested in nature and environment. Considering taking photos of historical sites or nature is a slight signal for their interests, we set the definiton of Nature label as in Table 1. People label, on the other hand, is for the posts from which the central interests seem to be on themselves. Medium label is for posts which are between People label and Nature label. The last label, Other label, is for the posts which are out of the contexts of the definitions of the other labels.

Table 1 The definition of four labels for focuses of SNS posts.

Labels	Definition	Example
Nature	Posts consist of images that focus on sigfhts, nature/plants, local cultures or history, and comments that describe them.	Fig. 2 a)
People	Posts consist of images focusing on making people look good or comments that do not describe sights, plants or nature, but describe their feeling at the site.	Fig. 2 b)
Medium	An intermediate label between Nature and People labels. Images and comments of posts with this label refer to people as well as sights, plants or nature.	Fig. 2 c)
Other	Posts are not related to any other labels, such as posts of foods, drinks, books and tableware.	Fig. 2 d)

3.2 Collecetd Data and Annotation

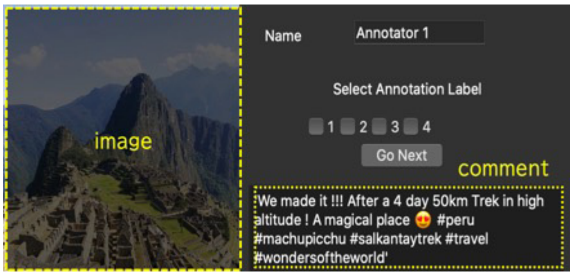
To judge users' focuses of their SNS posts, we hired six graduate students to annotate Instagram posts with our defined labels. We use Instagram posts taken in tourist sites which have various environmental issues caused by overtourism. To collect such posts satisfying the requirement, we selected six tourist sites suffering from overtourism problems, according to a Japanese version of Wikipedia article of overtourism . Three sites are in Japan: Biei, which is a popular ski area, Mt. Fuji and Shirakawa-gou, which are a World Heritage site. The other three sites, Easter islands famous for the statues Moai, Machu Picchu, which are famous ruins of the Inca Empire in Peru and Venice in Italy were selected. We gathered 500 posts for each site from Instagram by searching with *hashtags* of the six sites. Hashtags are special keywords starting with a # symbol for categorizing posts and making users find posts on particular topics. The used hashtags were #biei, #mtfuji, #shirakawa-gou, #easterislands, #machupicchu and #venice.

In the annotation process, we allocated three annotators for each site. The annotators were asked to put labels on posts according to the definition in Table 1 and examples with these labels. **Figure 4** displays a screenshot of the annotation software we developed. This tool shows the image and comments of an Instagram post in a window. With the tool, annotators select the corresponding labels for a post.

When the annotators faced difficult posts for annotations, we

---

[*2] https://www.tripadvisor.com/



Name: Annotator 1

'We made it !!! After a 4 day 50km Trek in high altitude ! A magical place 😍 #peru #machupicchu #salkantaytrek #travel #wondersoftheworld'

Fig. 4 A screenshot of our developed annotation software.

asked them to judge based on nature interest of the user who uploaded it. From our prior trials, we noticed that some cases of judging between Nature label and Medium label are difficult. In this case, we instructed that when the background of the post depicts landscapes about the site, then the post should be labeled with Nature label, and otherwise, the post should be labeled with Medium label. After the annotators did several trials of the annotation with the instructor, they started to annotate. We adopted multi-labels annotation because this annotation quite depends on one's subjectivity, and we intended to take the aspect in our created dataset.

3.3 Evaluation of the Dataset

We evaluated the outcomes of the annotation by checking the inter-annotator agreement in two ways: partial agreement by two annotators on a site (2A setting) and full agreement by three annotators on a site (3A setting). We set the two criterions because the annotation would be influenced by the annotators' subjectivities, and only 3A setting might be too strict for reflecting such judgment in our created dataset. We evaluated the inter-annotator agreement by dividing the number of posts with agreements by the number of all the posts for a site. We also computed the Kappa coefficient [5] as an indicator for evaluating the agreement.

Table 2 shows the evaluations and statistics of the created dataset. In the table, the inter-annotator agreement on 2A are high over 0.9 for all the tourist sites while the figures on 3A are low, at 0.292 with the worst case in Mt. Fuji. In terms of the Kappa coefficient, the value of Shirakawa-gou is over 0.8. Those of Biei, Easter islands and Venice are over 0.6. On these four cities, the Kappa values indicate a substantial agreement. The other two sites indicate a moderate agreement. The number of posts in the dataset is over 450 on 2A and that of 3A is below 350 for each site.

Figure 5 shows the distributions of the annotated labels on both settings. In both of the figures, the posts with Nature labels dominate a large part of the dataset. Compared to Nature label, the number of posts with other labels is small. In Fig. 5 B), the number of posts with agreements is generally small compared to those in Fig. 5 A) except for Nature label. That is, the posts on 3A setting represent our label definition in Table 1 clearer than the post on 2A setting.

We also investigate statistics of comments of Instagram posts, show summaries in **Table 3**. The table displays the average number of words (length of comments) and the size of used vocabularies in comments analyzed by using Stanford tokenizer implemented in python NLTK library [*3] for comments of Easter islands, Machu Picchu Venice (places over the world), and using MeCab Japanese morphological analyzer [*4] for those of Biei Mt. Fuji and Shirakawa-gou (Japanese places). The average number of words for a comment is roughly 82 words in Japanese places. That of places over the world is about 54.

Table 2 The statistics of the annotation's outcomes. IAA stands for inter-annotator agreement with two annotators (2A) and three annotators (3A).

Statistic	Biei	Mt. Fuji	Shirakawa-gou	Easter islands	Machu Picchu	Venice
IAA (2A)	0.970	0.896	0.980	0.964	0.910	0.962
IAA (3A)	0.694	0.292	0.638	0.580	0.406	0.618
kappa	0.619	0.321	0.808	0.622	0.455	0.649
# of posts (2A)	485	448	490	482	455	481
# of posts (3A)	347	146	319	290	203	309

Table 3 The statistics of comment parts of Instagram posts for two agreement settings of 2A and 3A.

Statistic	Biei	Mt. Fuji	Shirakawa-gou	Easter islands	Machu Picchu	Venice
Avg. # of words (2A)	90.1	94.3	82.0	56.1	62.0	45.4
Avg. # of words (3A)	65.6	91.8	66.6	56.0	65.7	36.8
Size of vocabulary (2A)	5,149	6,845	4,718	6,143	5,765	5,937
Size of vocabulary (3A)	1,661	2,733	1,750	2,535	1,375	1,435

The size of vocabularies is about 5,500 for comments of Japanese places and 5,900 for those in places over the world on 2A setting. The size of vocabulary on 3A is smaller than that of 2A because of the small number of posts in 3A. We did not distinguish keywords in hashtags and ordinal words in comments. In Instagram, hashtags play a different role from ordinal words in users' comments, and should be treated as different symbols. However, the number of posts in a single site is limited, about 350 posts in Biei on 2A at the max. The number of hashtags in one site is small as well. In addition, some users only write every single words in the hashtag fashion in their comments such as ordinal comments. That is why we treat hashtags as ordinal words by removing # for this study.

4. Experiments

Because no studies have addressed to the DFPT task to the best of our knowledge, the purpose of this experiment is to evaluate how much the task can be solved by the recent classification methods, which can be seen as the baselines. In this section, we describe the experimental setup, and present the results. We then analyze successful cases and failures of the classification.

[*3] http://www.nltk.org/api/nltk.tokenize.html

[*4] https://taku910.github.io/mecab/

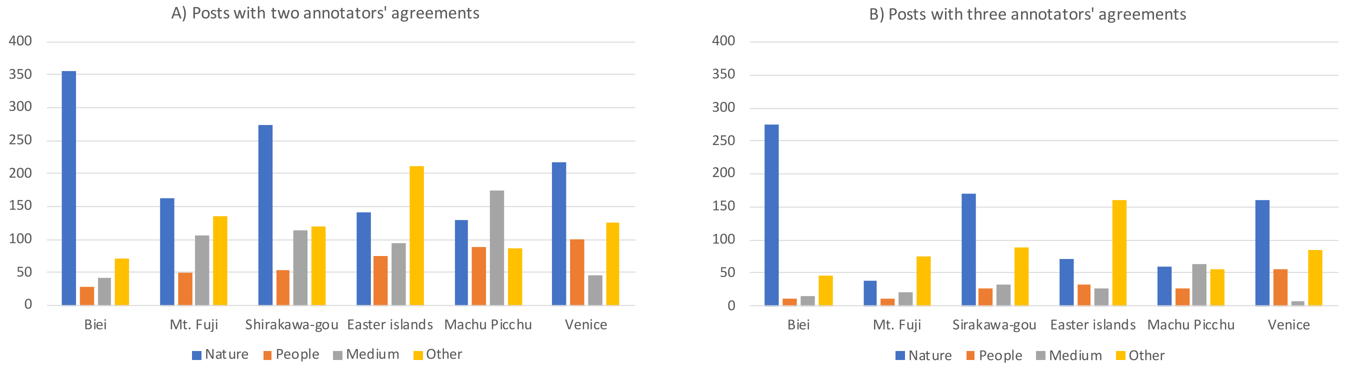


Fig. 5 The number of posts agreed by A) two or B) three annotators on each label for the six tourist sites.

4.1 Experimental Setup

In the experiment, our created dataset is used for the evaluation in two settings derived from the annotation in Section 3. The first (resp. second) setting uses correct labels with which two (resp. three) annotators agreed. The first (resp. second) one is called 2A (resp. 3A) setting. We evaluate classification methods based on two policies. One policy uses features of pictures of Instagram posts, which we call an image-based method. The other policy utilizes users' comments of Instagram posts, which we call a text-based method. The yellow regions of Fig. 4 stand for examples of the used parts of Instagram posts for the two features.

First, to implement image-based methods, we employed three CNN models, the Alexnet [12], the VGG net [19] and the ResNet [6]. We used the three models pretrained on more than one million images in the ImageNet [3] database for 1,000 objective classes. This may be beneficial for classifying "Other" label in the DFPT since the class system includes labels related to scenes of eating such as "restaurant", "carbonara" and so on, which sometimes appear in the photos of "Other". In terms of setting of the AlexNet, the output layer of the model is replaced to the size of four labels that we define. We resized input images to 227 × 227 to be fed into the model. For settings of the others, we used the 19-layer models of VGG and the 50-layer models of the ResNet, and added a fully-connected layer of four unit size onto the top of both models with the softmax activation function. The three models were optimized with stochastic gradient descent (SGD), with a maximum epoch of 30 and a batch size of 16.

Second, we selected the *Long Short Term Memory* (LSTM) [7] and the *Bidirectional Encoder Representations from Transformers* (BERT) [4] for the text-based classification models. The LSTM is often used as a baseline method for text classification tasks. The BERT is a recent language model based on transformers. It is often used for a variety of NLP tasks by fine-tuning the model. The LSTM model consists of six layers: an input layer, a word embedding layer, an LSTM layer, a fully-connected layer, a Softmax layer and a classification layer. We set the size of the word embedding layer at 100 and the number of hidden units in the LSTM layer at 180. The used tokenizers are MeCab morphological analyzer for Japanese comments and Stanford tokenizer for English comments as described in Section 3.3. (From here, tokenizers used in the BERT are same to these.) The maximum length of words for the input was 200 because the average lengths of words is less than 100. This model was optimized with the Adam optimizer, with a maximum epoch of 30 and a batch size of 16.

For a BERT classification method, two pretrained BERT models were used for Japanese comment classifier and English comment classifier. The architecture of both models consists of 12 layers, 768 dimensions of hidden states and 12 attentions heads. The Japanese model is trained on dump data of Japanese Wikipedia articles for a masked language modeling objective [4] with the whole word masking manner in which all of the subword tokens corresponding to a single word are masked at once. The English model is pretrained on BookCorpus [23], which is a corpus of unpublished books, and English Wikipedia articles for a masked language modeling objective.

From the two pretrained models, we fine-tuned the models for comment classification of the DFPT. The token size for models is 512 tokens with padding for comments less than that size and truncating those over that size. The loss function was cross-entropy of the classification. This model was optimized with the Adam optimizer, with a maximum epoch of 30 and a batch size of eight. To implement the above BERT classification methods, we used the BERT pretrained models opened in HuggingFace Pretrained models [*5]. Concretely, the used models were 'cl-tohoku/bert-base-japanese-whole-word-masking' for Japanese and 'bert-base-cased' for Englsih in the site.

Evaluations were conducted with holdout testing dividing the posts of the dataset into 70% for training, 15% for validation and 15% for testing.

4.2 Results

Tables 4 and **5** show accuracies of the classification methods. The image classifier of the VGG19 achieves the best accuracy on both tables showing the macro averaged accuracy of 0.840 on 3A and 0.789 on 2A. The second best method is the comment classifier of the BERT showing the macro averaged accuracy of 0.795 on 3A and 0.622 on 2A. In terms of tourist sites, the VGG19 shows the best results in all of the sites except for Easter islands and Machu Picchu, in which the BERT shows the best accuracy, in Table 4.

Next, we compare performances in the two tables to evaluate the setting for agreement in the annotation. The figures are gen-

---

[*5] https://huggingface.co/transformers/pretrained_models.html

Table 4 Accuracy of the two classifiers on 3A setting.

Method	Biei	Mt. Fuji	Shirakawa-gou	Easter islands	Machu Picchu	Venice
Image (AlexNet)	0.904	0.571	0.745	0.744	0.800	0.761
Image (ResNet)	0.810	0.686	0.614	0.630	0.542	0.618
Image (VGG19)	**0.931**	**0.875**	**0.825**	0.722	0.813	**0.872**
Comment (LSTM)	0.789	0.524	0.617	0.558	0.433	0.544
Comment (BERT)	0.842	0.667	0.742	**0.930**	**0.867**	0.719

Table 5 Accuracy of the two classifiers on 2A setting.

Method	Biei	Mt. Fuji	Shirakawa-gou	Easter islands	Machu Picchu	Venice
Image (AlexNet)	0.743	0.522	0.607	0.590	0.521	0.726
Image (ResNet)	0.622	0.521	0.562	0.5	0.444	0.667
Image (VGG19)	**0.824**	**0.704**	**0.825**	**0.727**	**0.892**	**0.764**
Comment (LSTM)	0.716	0.612	0.518	0.449	0.197	0.507
Comment (BERT)	0.623	0.625	0.631	0.718	0.535	0.600

erally high on 3A setting in comparison to those on 2A setting. This might be because the criterion of agreement by three persons (3A setting) is more strict than two persons' agreement. This possibly results in that posts agreed by three annotators would have obvious clues in the visual or the text written in comments.

Since the image classifier (the VGG19 method) shows the best accuracy among the examined methods, we analyze performances of the method on each label by evaluating the F-measure to see if there exists differences of performance among labels. The precision (P) and the recall (R) are defined as follows:

$$\text{Precision} = \frac{\#\text{ of the posts classified correctly}}{\#\text{ of the posts classified into the label}},$$

$$\text{Recall} = \frac{\#\text{ of the posts classified correctly}}{\#\text{ of the posts of the label in the dataset}}.$$

The F-measure is the harmonic means of the recall and the precision.

Table 6 shows values of the F-measure (F) on 3A setting, and **Table 7** shows those on 2A setting. The relatively high values are constantly observed in only Nature and Other labels over all the sites in Table 6. The values of Medium and People labels are low indicating 0 in Mt. Fuji and Shirakawa-gou. This might be due to the lack of posts for both training and testing as we investigated in Fig. 5. The number of testing posts for sites, which show 0 of the F-measure in People and Medium labels, were only two to four. By contrast, because there were several posts of People and Medium labels in 2A, the F-measure of the two labels are relatively high in Table 7.

We also observed that the F-measure of all the four labels on Machu Picchu is relatively well compared to other places in both of 3A and 2A. One of the reason may be the balanced number of posts for each label as shown in Fig. 5.

Table 6 F-measure of the image classifier on 3A setting.

Label	Biei	Mt. Fuji	Shirakawa-gou	Easter islands	Machu Picchu	Venice
Nature	0.959	0.955	0.944	0.732	0.882	0.941
People	0.667	0	0	0	0.884	0
Medium	0	0	0	0.25	0.4	0.846
Other	0.857	0.769	0.733	0.836	0.571	0.571

Table 7 F-measure of the image classifier on 2A setting.

Label	Biei	Mt. Fuji	Shirakawa-gou	Easter islands	Machu Picchu	Venice
Nature	0.920	0.841	0.825	0.727	0.891	0.764
People	0.286	0.462	0.444	0.4	0.923	0.143
Medium	0	0.25	0.526	0.375	1	0.717
Other	0.762	0.546	0.737	0.839	0.882	0.636

4.3 Analysis of Results

We experimented with image classification methods and comment classification methods on 2A and 3A setting respectively. The highest classification accuracy is achieved by the image classification method of the VGG19. Therefore, we analyze the correctly classified and miss-classified posts by the image classification method. **Figures 6** and **7** show correctly classified examples and miss-classified examples on 3A setting, respectively. The example images in Fig. 6 seem to suit the label definition well though they were randomly sampled. In other words, posts conforming the label definition well with strict agreements of three annotators can be classified well.

While, posts, which are difficult for even humans to distinguish their focuses, tend to be misclassified. We investigate error cases of Nature and People labels. First, on Nature label, the classification was failed on the posts from which you can't determine the tourist sites easily. Figure 6 a) displays pictures depicting the famous spots in the tourist sites, for example, Machu Picchu, the houses with traditional roofs for Shirakawa-gou and the famous blue-pond in Biei. For these pictures, you can determine the sites from the images. However, the miss-classified cases in Fig. 7 a) are difficult to distinguish the site from their pictures. In Fig. 7 a) 2), for example, the Moai statue on Easter island is captured, but it might be difficult to judge that because of the sunlight. As you can see from these classification examples, the classifier learned not whether a place is about nature or not, but whether it is a characteristic spot of the tourist site.

Second, on People label, the miss-classification was caused by subject mixtures of people and other stuffs. For the successful examples of the label, Fig. 6 b) depicts people as their main focuses. By contrast, for example, Fig. 7 b) 1) might be taken for the person, the plants and the row of houses, and Fig. 7 b) 2) contains the area of the people and that of the airplane in the picture.

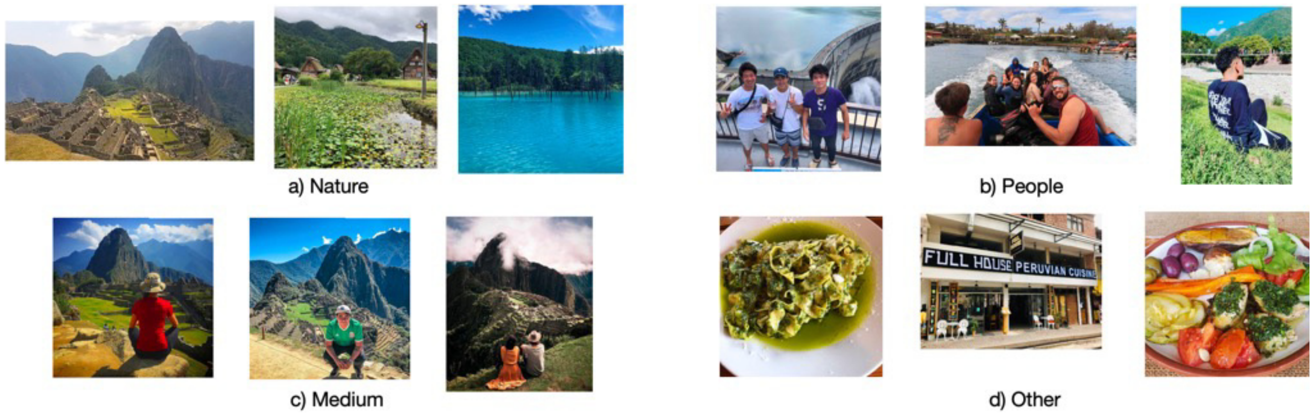


Fig. 6 Examples of correctly classified posts.

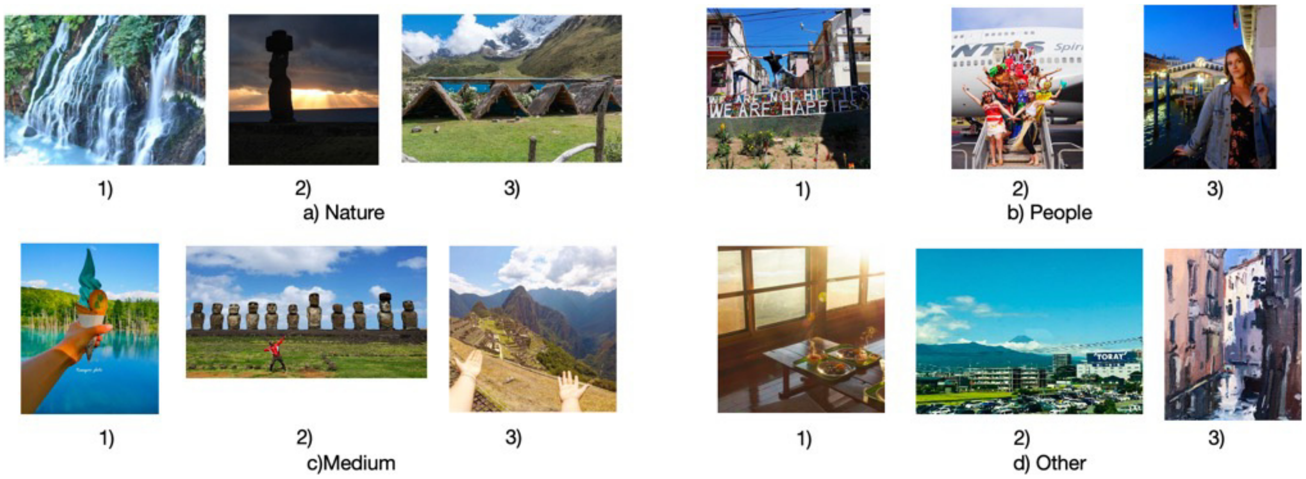


Fig. 7 Examples of miss-classified posts.

4.4 Discussion

In the experiments, we evaluated image classifiers based on visual features of Instagram posts or text classifiers based on the comments in the posts separately. The main reason for evaluating classifiers using one type of data in the posts is to investigate which part of the image or the comment in a post can contribute to classifying the focuses of the user's post for this time.

However, classification methods using the combination of features of images and texts show a higher performance than methods relying on only one type of data in the image classification task [11], [21] or in the sentiment analysis of images on SNS [10]. For an example in these studies, Kiela *et al*. proposed an image classification method based on concatenation of feature vectors extracted from a trained CNN image classifier and word embeddings of tags attached to images [11]. Their method improves the accuracy of the image classification task by about 0.05 at the max compared with classification methods using solo feature.

The performance of the DFPT is possibly increased by such classifiers based on combined features of both images and comments in the posts. In Instagram posts, over 80% of posts contain at least one hashtag in the comment in our dataset. Some hashtags denote objects, places in the photos, which are close to tags in Ref. [11]. We also observed that emojis are used to express users' feelings richly in about half the posts. The combination feature of both comments and images is expected to learn useful relations about users' subjectivities hidden in photos, hashtags or emojis for the DFPT task.

5. Questionnaire Investigation

In order to examine connections between our defined focuses and users' interest for environments, we conducted a questionnaire survey about that. The purpose of this investigation specifically is to answer the following questions:

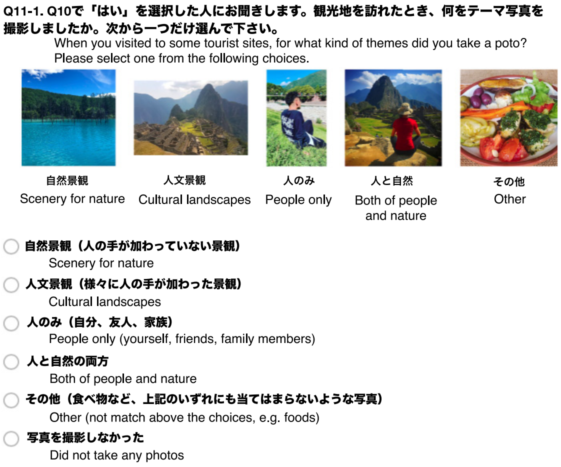
(1) Are there people matching our expected potentially eco-friendly people?
(2) Is it true that focuses of images in SNS posts reflect a user's interest in environments?
(3) If the second question is yes, then can the DFPT framework and the developed classifiers actually grasp a difference of focuses of images among interest groups using real-world data of Instagram posts?

In this section, we explain questions of the questionnaire and how to conduct the survey. We then explain the outcomes of the survey.

The questionnaire survey was conducted in a form of an online questionnaire using Yahoo! Japan crowd sourcing service [*6]. We prepared a total of 12 questions for the questionnaire form. Each of the questions is about interests of environmental issues at tourist sites or questionee's behavior of sharing his/her tourism

---

[*6] https://crowdsourcing.yahoo.co.jp/



Q11-1. Q10で「はい」を選択した人にお聞きします。観光地を訪れたとき、何をテーマ写真を撮影しましたか。次から一つだけ選んで下さい。
When you visited to some tourist sites, for what kind of themes did you take a poto?
Please select one from the following choices.

○ **自然景観（人の手が加わっていない景観）**
Scenery for nature
○ **人文景観（様々に人の手が加わった景観）**
Cultural landscapes
○ **人のみ（自分、友人、家族）**
People only (yourself, friends, family members)
○ **人と自然の両方**
Both of people and nature
○ **その他（食べ物など、上記のいずれにも当てはまらないような写真）**
Other (not match above the choices, e.g. foods)
○ **写真を撮影しなかった**
Did not take any photos

Fig. 8 Screenshot of a question in the online questionnaire. English subtitles are arranged for this paper and were not displayed for the participants.

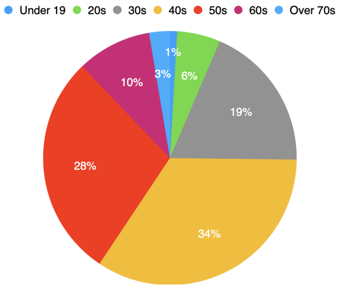


Fig. 9 The proportion of questionees' ages.

experience on SNS. One example of question is displayed in **Fig. 8**, which is a question about focuses of photos when a questionee visits tourist sites.

We collected 2,838 answers in all from the online questionnaire. After omitting 95 invalid answers, such as selecting a same answer choice for all the questions, the number of effective answers is 2,743. The percentage of male questionees is 63.7%, that of female is 35.1% and that of persons without an answer of their sex is 1.2%. The statistics on the participants' age are summarized in the pie chart of **Fig. 9**. In the figure, most of the participants are in their 40s at 34%. The second and third largest proportions of participants are in their 50s and in their 30s at 28% and 19%, respectively. In Japan, most Instagram users are young people such as teenagers or in their 20s according to the report of Japanese Ministry of Internal Affairs and Communications [8]. The distribution of participants' ages for this survey is different from that of Instagram users. However, this questionnaire reaches people in most generations, which is desirable for our goal of predicting people's awareness for environments across all ages.

5.1 Cross Tabular Analysis for Confirming Potentially Eco-friendly People

To answer the question (1) "Are there people matching our expected potentially eco-friendly people?", we investigated that there are such people by asking questionees' current interest for environmental isssues and their future interest for it in the questionnaire. We asked following two questions:

- Question 8-1: "Please select the level of interest for environmental protection for tourist sites, for example, avoiding littering, preventing environmental destruction or protecting ecosystems of sites". Answer choices are 1) strongly interested, 2) rather interested, 3) not rather interested or 4) not interested.
- Question 8-2: "Please select the level of future interest for environmental protection for tourist sites, for example, avoiding littering, preventing environmental destruction or protecting ecosystems of sites". Answer choices are you will be 1) strongly interested, 2) rather interested, 3) not rather interested or 4) not interested.

Table 8 Cross tabulation results of Question 8-1 (Current interests) and Question 8-2 (Future interests). Figures stands for the number of questionees.

Q8-1 / Q8-2	Choice 1	Choice 2	Choice 3	Choice 4
Choice 1	208	23	2	0
Choice 2	75	1,302	61	5
Choice 3	5	320	652	13
Choice 4	2	7	39	139

We performed a cross-tabulation analysis of the answers for the two questions. The results are reported in **Table 8**. The number of people who may have stronger interest for environmental issues in the future than they do is, for example, 75 at (Choice 2, Choice 1), 5 at (Choice 3, Choice 2) and 320 at (Choice 3, Choice 2) in the table. The sum of the number of people have a storng interest in the future is higher than that of people have a weak interest in the future.

5.2 Statistical Test between Focuses of SNS Photos and Users' Awareness for Environments

To answer the question (2) "Is it true that focuses of images in SNS posts reflect a user's interest in environments?", we verify whether there exists a statistical significance between the focuses of SNS photos and users' interests with statistical testing. Thus, we set up a null hypothesis that there does not exist a statistical significance between the focuses of SNS photos and users' interests.

First, we divided questionees into two groups, people with some experience of visiting some tourist sites or people without such experience, because activities about taking photos may be different among the two groups influenced by experiences. The number of people with travel experience is 2,411, and that of people without such experience is 332. Second, we grouped answer choices of Question 11 and 12 into two classes of interests. Because taking photos of particular scenery of nature or architectures related to tourist sites may be sign of their potential interest for environments, we sorted choices 1) and 2) into class A) and the other choices 3) – 5) into class B). The asked questions 11 and 12 are as follows:

- Question 11: "What kind of themes did you take photos for when you visit a tourist site? Please select one theme from answer choices". Answer choices are 1) Scenery of nature, 2) Scenery of architectures, 3) People only, 4) Both people and nature, 5) Other or 6) Did not take any photos. Example pictures of themes were displayed for questions, as shown in Fig. 8.

- Question 12: “What kind of themes do you take photos for if you visit a tourist site? Please select one theme from answer choices”. Answer choices are same to Question 11.

We verified a difference of the above two classes, class A) and B), and levels of interests for environments with one-sided t-testing at one percent level of significance. The testing results shows that the p-value is 5.23×10^{-7} for people with travel experience and 8.56×10^{-3} for people without travel experience. Since the p-value on both settings are lower than 0.01, the null hypothesis is discarded. Thus, we conclude that there exists a statistical significance between levels of people’s interests for environments and focuses of taking photos (class A and class B).

5.3 Investigation of the Focuses of Images on Each Interest Level Using Real-world Data

The purpose of this investigation is to answer the question (3) “If the second question is yes, then can the DFPT framework, the defined focuses of posts and the developed classifiers, actually grasp a difference of focuses of images among interest groups using real-world data of Instagram posts?” In this section, we collect all the posts of users who provided their Instagram user ID in the questionnaire. We then detect focuses of posts using classifiers examined in Section 4, and calculate the proportions of focus for each level of environmental awareness. After that, we investigate the relation between the levels and the proportions of focuses.

First, we collected Instagram posts of participants who provided their user ID for this investigation. Although 77 participants agreed to provide their accounts for this survey, we used posts of 26 users, who uploaded more than 10 posts and less than 1,000 posts, because the rest of users uploaded no posts or only a few posts. We collected a total of 4,914 posts, 189 on average for each user.

Second, as to classifiers for the focuses of posts, we used the image classifier of the VGG19 trained on all the pictures of the six tourists sites with 2A setting in Section 4.1 (results corresponded to Table 5 and Table 7), since the VGG19 classification method shows the highest F-measure over the four focuses labels in the six cities.

The results of the classification are displayed in **Fig. 10**. A bar-chart shows the proportion of classified focuses of photos for each level of interest for environments, which corresponds to each answer choice of Question 8-1 in Section 5.2. Overall, the proportion of “Other” label is large in all interest levels. This is because proportions of pictures such as foods or clothes are large in the collected posts. Aside from “Other” label, the proportion of “People” increases from 4.0% at “Interested in” to 13.0% at “Not interested in” as the level of interest for environments deteriorates. By contrast, that of “Nature” decreases from 24.5% at “Interested in” to 5.3% at “Not rather interested in”.

However, posts of “Not interested in” contain a large proportion of “Nature” label at 21.5%. The result of “Not interested in” group is possibly due to the biased sampling of users. The number of users and the total number of posts is three users and 851 posts in “Not interested in” while that of “Rather interested in” is 13 users and 1,486 posts. The number average posts for each user of “Not Interested in” is third times greater than that of “Rather interested in”.

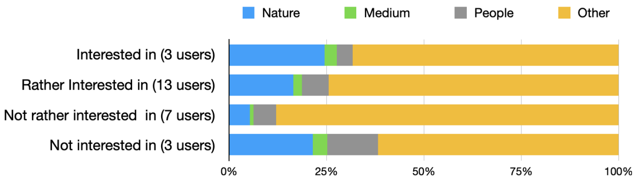


Fig. 10 Proportion of classified labels of focuses for each level of interest.

Table 9 Statistics and classification results when grouped into binary groups of interests.

Interest	# of users	# of posts
Interested in	16	2424
Not interested in	10	2450

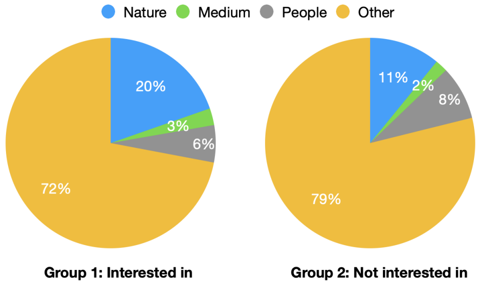


Fig. 11 Proportion of posts for each level of interest in environments.

Splitting the four levels of interest into two groups, we further investigated the relation between the proportion and the interest because the number of photos is not balanced for each level. Specifically, we sorted “Interested in” and “Rather interested in” into “Interested in” group. As statistics are shown in **Table 9**, the number of posts for “Interested in” users is about the same as that of the other one.

The results of the investigation are shown in **Fig. 11**. The proportion of posts of “Nature” is 20% for the group with the interest and 11% for the group without the interest. There is a difference of the proportion by 9% among the two, and this probably reflects the interest for environments influencing users’ choice of uploading photos to Instagram. Next, the proportion of “People” and “Other” for “Interested in” group is 6% and 72%, respectively. The proportion of the same labels for “Not interested in” group is higher than that of the other group at 8% and 79%, respectively. That is, the users who are not interested in environments tend to share photos with focuses of people or foods, which are not related to nature.

5.4 Discussion

In this section, we conducted a questionnaire investigation to analyze focuses of SNS posts and users’ awareness. We confirmed that there is a statistical significance between levels of people’s interests for environments and focuses of photos uploaded to Instagram.

We also confirmed that proportions of post focuses in users’ posting history are different from that of users with a different level of environmental awareness with a real-world data analysis. Again, the goal of this study was to find potentially eco-friendly people, and we investigated the existence of potentially eco-friendly people as shown in Table 8. But, there is a still gap between the proportion of users' focuses identified by the image classifier and their future interest. That remains for future research.

We note three shortcomings of this investigation setting. The first point is that we could only use the very small number of the real-world data in the analysis. Only 26 users' data were available, even 77 participants out of about 2,700 participants provided us their user ID, because of the essential difficulty of collecting such personal data, and we just asked voluntary participants in the questionnaire. Although we used about 4,900 posts in all, which would be not a few, the number of real-users should be increased by, for example, hiring SNS users directly.

The second point is about the label definitions of the DFPT. We used the classification labels for focuses of posts defined in Section 3.1. The labels are judged by third persons in the annotation process. Thus, the focuses classified by the trained classifiers would not fully reflect the intentions of original users' posting. This may reflect the results of a large proportion of Nature labels on "Not Interested in" group, which may take photos just for fun. One of the solution for this problem would be arranging a more fine-grained label system for detecting focuses.

The third point is about classification methods. We used the image classifier of the VGG19 with about 0.84 of accuracy on 3A setting evaluated in Section 4.2. The current performance of the classification is not very high, and miss-classification results must be included in the real-world data analysis. In that case, the relation between users' awareness and focuses of posts are misleaded to some extent. Improving classification methods is an important future work. Next, the image classifier was trained on a mixture of the six cities selected in Section 3.2, which are different locations, consist of a half of Japanese tourist sites. It would be good to check a performance by using same locations both in training data and testing data.

6. Conclusions

In this paper, we studied two research topics to grasp people's awareness for environments from posts of SNS. The first topic is to setup a research task to classify posts of SNS sites into their awareness for nature. The second topic is to analyze the relation between such awareness and their posts shared on an SNS site with a questionnaire investigation.

First, we formalized a research task called Detecting Focuses of Posts about Tourism (DFPT), in which a system classifies SNS posts about tourist sites into four types of focuses of their posts. To lay a basis of evaluating the task, we created an evaluation dataset. We then evaluated the task with simple classification methods: image classification methods with CNNs and comment classification methods with the LSTM or the BERT. Experimental results show that the image classifier using the VGG19 achieves the highest performance at about 0.84 of accuracy in the examined methods on the three-annotators agreement setting. That is a baseline result for future systems.

Second, we statistically analyzed people's awareness for environments and the relation between their awareness and their SNS posts with a questionnaire investigation covering over 2,700 people. The analysis results show that there is a statistical significance between levels of people's awareness for environments and the focus types of SNS posts. Furthermore, with the investigation of the real-world data of Instagram, we confirmed that there are differences between users' interest and proportions of the focuses of posts in their posting history, which can be a clue for estimating people's awareness for next research steps.

For future work, the improvement of classification methods is necessary. In the experiments, the classification performance for labels "Medium" and "People" is low due to the shortage of the number of posts for some sites in the training data, such as Mt. Fuji and Shirakawa-gou. It is possible to improve it by increasing the number of posts in the dataset. We also note the low performance of the F-measure on "People" label and "Medium" label, arising from the fact that it is intrinsically difficult to distinguish them. We need to devise a classification method for distinguishing the two labels, such as taking visual features of the area of people into the method. In terms of analysis with the real-world data of Instagram, the number of users was very limited this time because collecting such personal data is difficult in nature. The analysis with additional data remains for future research.

References

[1] Baba, T., Baba, K. and Ikeda, D.: Detecting Mental Health Illness Using Short Comments, *Proc. 33rd Advanced Information Networking and Applications*, pp.265–271 (2019).

[2] Centola, D., Willer, R. and Macy, M.: The Emperor's Dilemma: A Computational Model of Self-Enforcing Norms, *AmJ. Sociol.*, Vol.110, No.4, pp.1009–1040 (2005).

[3] Deng, J. et al.: ImageNet: A Large-scale Hierarchical Image Database, *Proc. 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp.248–255 (2009).

[4] Devlin, J., Chang, M., Lee, K. and Toutanova, K.: Bert:pre-training of deep bidirectional transformers for language understanding, *Proc. 2019 Conference on the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp.4171–4186 (2019).

[5] Fleiss, J.L., Levin, B. and Paik, M.C.: The Measurement of interrater agreement, Shewart, W.A. and Wiks, S.S. (Eds.), *Statistical Methods for Rates and Proportions, Second Edition*, pp.212–236 (1981).

[6] He, K. et al.: Deep residual learning for image recognition, *Proc. 29th IEEE Conference on Computer Vision and Pattern Recognition*, pp.770–778 (2016).

[7] Hochreiter, S.: Long Short-Term Memory. *Neural Comput.*, Vol.9, No.8, pp.1735–1780 (1997).

[8] Institute for Information and Communication Policy, Japanese Ministry of Internal Affairs and Communications: Report on Usage of Information Commnication Media and Activities on Information in 2020 (2020) (in Japanese), available from ⟨https://www.soumu.go.jp/main_content/000708016.pdf⟩ (accessed 2021-03-01).

[9] Kashiwagi, N., Suzuki, T., Lee, J. and Ikeda, D.: Classification of multimedia sns posts about tourist sites based on their focus toward predicting eco-friendly users, *Proc. 2nd ACM Multimedia Asia*, No,19, pp.1–7 (2021).

[10] Katsuri, M. and Satoh, S.: Image sentiment analysis using latent correlations among visual, textual, and sentiment views, *Proc. 41st IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.2837–2841 (2016).

[11] Kiela, D. et al.: Efficient large-scale multi-modal classification, *Proce. 32nd AAAI Conference on Artificial Intelligence*, pp.5198–5204 (2018).

[12] Krizhevsky, A., Sutskever, I. and Hinton, G.E.: Imagenet classification with deep convolutional neural networks, *Proc. 26th Conference on Neural Information Processing Systems 2012*, pp.1097–1105 (2012).

[13] Leibowitz, A.: Overtourism has travel advisors telling their customers to please go somewhere enlse, *Skift* (2019), available from ⟨https://skift.com/2019/03/05/overtourism-has-travel-advisors-telling-their-customers-to-please-go-somewhere-else/⟩ (accessed 2019-09-15).

[14] Li, S. et al.: Analysis of sns photo data taken by foreign tourists to japan and a proposed adaptive tourism recommendation system, *Proc.*

2017 International Conference on Progress in Informatics and Computing, pp.323–327 (2017).
[15] Liu, P., Guberman, J., Hemphill, L. and Culotta, A.: Forecasting the presence and intensity of hostility on instagram using linguistic and social features, *Proc. 12th International AAAI Conference on Web and Social Media*, pp.181–190 (2018).
[16] Min-ho, J.: Jeju suffers from 'too many tourists', *The Korea Times* (2018), available from ⟨https://www.koreatimes.co.kr/www/news/nation/2018/01/371_241832.html⟩ (accessed 2019-09-15).
[17] Ofli, F. et al.: Is saki #delicious?: the food perception gap on instagram and its relation to health, *Proc. 26th International World Wide Web Conference*, pp.509–518 (2017).
[18] Ohkubo, R. and Muromach, Y.: A study of destination images of foreign tourists to japan by analyzing travel guidebook and review site, *J. City Plan. Inst. Japan*, Vol.49, No.3, pp.573–578 (2014).
[19] Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, arXiv preprint, arXiv:1409.1556 (2014).
[20] Suzuki, S.: Analysis of covert and overt interest in tourist attractions with use of twitter data: Use of social media data in destination marketing, Ph.D. Diss., Tokyo Metropolitan University (2017).
[21] Wang, G., Hoiem, D. and Forsyth, D.: Building text features for object image classification, *Proc. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp.1367–1374 (2009).
[22] Xi, N. et al.: Understanding the political ideology of legislators from social media images, *Proc. 14th International AAAI Conference on Web and Social Media*, pp.726–737 (2020).
[23] Zhu, Y. et al.: Aligning books and movies: towards story-like visual explanations by watching movies and reading books, arXiv preprint, arXiv:1506.06724 (2015).

Tokinori Suzuki is currently a project-specific assistant professor in the Department of Informatics, Kyushu University. He received his Bachelor of Informatics, M.E. in Computer Science and Ph.D. in Informatics from Senshu University in 2012, and Tokyo Institute of Technology in 2014 and Kyushu University in 2021, respectively. His research interest includes information access and web mining. He is a member of the Information Processing Society of Japan (IPSJ).

Naoto Kashiwagi is a business consultant at Deloitte Tohmatsu Consulting LLC. He received his Bachelor of Information Technology and M.E. in Informatics from Kyushu Institute of Technology in 2019 and Kyushu University in 2021, respectively.

Jounghun Lee is a visiting researcher in Environmental Planning Institute, Seoul National University, Korea. She received her Ph.D. in Science from Kyushu University, Japan. Her main research interest is sustainable management of common resources and the social dilemma that undermines the successful management of focal resource. Simplified theoretical model (evolutionary game theory) and simulation are used to study the problem.

Kun Qian is currently a lecturer at Faculty of Humanities, Fukuoka University. He received his Ph.D. of psychology from Kyushu University in 2012. He worked in Kyushu University as an assistant professor (2014–2021) and in ABeam Consulting Ltd. (2012–2014) as a certified SAP consultant. His current research concerns cultural and experimental psychology. He is a director of Kyushu Open University, and a member of the Japanese Psychological Association.

Daisuke Ikeda is currently an Associate Professor in the Department of Informatics, Kyushu University. He received his bachelor degree, master degree, and doctor degree from Kyushu University in 1994, 1996, and 2004, respectively. His research interests include data analysis and data infrastructure. He is a member of IPSJ and ACM.