Exploring the Relationship between User Activities and Profile Images on Twitter through Machine Learning Techniques

Tomu Tominaga[1] and Yoshinori Hijikata[2]

[1]*Graduate School of Engineering Science, Osaka University, tominaga@hlab.sys.es.osaka-u.ac.jp*

[2]*School of Business Administration, Kwansei Gakuin University, contact@soc-research.org*

ABSTRACT

Social media profile images are one of many visual components of users. Moreover, user activities such as posting or chatting are regarded as self-expression behaviors. In this study, we examine Japanese Twitter users to explore the relationship between user activities and profile images. Logistic regression analysis is used to statistically identify and quantify relationships, leading us to conclude that several profile image categories significantly correlate with user activities. Furthermore, we use machine learning techniques (logistic regression, random forest, and support vector machine) to predict whether or not a user belongs to a specific profile image category. Each model's performance is evaluated and compared for all profile image categories. Primary results show that users whose profile image includes others' faces are more likely to use a replying function but less likely to add url links to their tweets, and that it is the easiest for machine learning models to find their category from their user activities. In short, our findings indicate that visual expression correlates with social media user behavior.

Keywords: Twitter, User activity, Profile image, Micro-blog

1 Introduction

People can deliver and receive information on social media. Generally, social media user actions related to information sharing is explicitly provided on a large scale. To date, researchers have examined these and have applied results to personalized services (Pan et al., 2013; Abel et al., 2011a; Hannon et al., 2010), event detection (Weng and Lee, 2011; Sakaki et al., 2010), or marketing analyses (Ikeda et al., 2013).

Twitter is a massive, worldwide social media presence. Compared to other social media, Twitter possesses two unique qualities: posting size is smaller, and people can connect and get updates without permission. On Twitter, posts are limited to 140 characters. Therefore, slang words and emojis are more likely, because users should briefly express opinions and feelings. Moreover, in contrast to Facebook or LinkedIn, Twitter allows users to follow others without their consent, making it easy for users to share information. Because of these unique factors, researchers conducted a Twitter-specific text analysis of user content (Abel et al., 2011b; Ramage et al., 2010; Zhao et al., 2011) and examined the sphere of information diffusion throughout user networks (Kwak et al., 2010; Pan et al., 2013).

We assume that not only texts in tweets and links of user networks but also images might also reflect users' characteristics. Although several types of images are used on Twitter such as post-attachments and user profiles, we focus on profile images (i.e., user pictures set on a profile page for identification). These images are displayed on Twitter timeline, where tweets from following users are chronologically streamed. These images play an important role in identifying tweet owners; they are visual proxies. Liu et al. (2016) and Hum et al. (2011) dis-

cussed that profile image choices correlate to various user attributes (e.g., psychology, personality, preferences, and motivations). We follow that user activities on Twitter are influenced by those user attributes, because activities can be regarded as user self-expression. For example, if a user prefers communicating with others, he or she will often use the Twitter "reply" function. Moreover, if a user wants to advertise content, he or she will likely attach URL links.

In this study, we analyze how profile image categories correlate to user activity types for Japanese Twitter users. Although not causally well-understood, we believe that there is a set of properties that correlates profile images to those activities. Findings from our examination can be applied to user personalization, where the system recommends profile images based on user activities. Findings can also be applied to marketing analyses, by determining user preferences from profile images. These applications possess the advantage of native language-independence, but, as discussed later, not necessarily cultural independence.

To explore the relationship between profile images and user activities, we manually sort profile images of Japanese Twitter users into 13 categories. This categorization is based on an objects that appear in the images. We then examine user activities per those profile image categories.

The present study provides three key contributions. Firstly, to the best of our knowledge, this is the first study to comprehensively explore the relationship between profile images and user activities on Twitter. Second, we find significant differences in user activities across several profile image categories. Third, we build prediction models to infer whether users belong to a specific profile images category from their user activities.

The rest of this paper is organized as follows. First, we review conventional work related to this study. Afterwards, we describe methods for data collection and analysis, and present our results, including implications. Finally we state limitations of our study and summarize this paper.

2 Related work

Profile images are one type of implicit visual expression for users (Goffman, 2002). Visual appearance is associated with psychological states of users (Fong and Mar, 2015; Turner and Hunt, 2014). Researchers have examined how, why, and where (e.g., social media or online dating sites) users use or select their profile images.

Zhao and Jiang (2011) assessed profile images used by university students in U.S. and China on Facebook and Renren. They concluded that U.S. students are more likely to use photos that include friends than Chinese students are. Other than Facebook, various categories of profile images were reported on Twitter (Tominaga and Hijikata, 2015) and on YouJust-GetMe (Steele Jr et al., 2009). Hum et al. (2011) conducted content analysis to understand how high school and college students appear in their Facebook profile images. They stated that most users set several profile images, in which they appear socially appropriate and not so active. In terms of impression management, Siibak (2010) investigated young male users' self-presentation through profile images on Rate, a popular social networking service (SNS) in Estonia. Their study reported that men aged 15 to 28 try to select profile images to express sexual and romantic contexts. Siibak and Hernwall (2011) conducted a survey observing how Swedish and Estonian teens construct gender identities through profile images on SNS. They noted that girls aged 12 to 16 modify their profile image to make it more attractive. Per a study by Kapidzic (2013), narcissism is significantly related to motives for selecting attractive and unique profile images. On Facebook, when communicating with others, users tend to prefer attractive profile images of the opposite gender (Wang et al., 2010). Moreover, females are more likely to change profile images to emphasize their social relationship status (Strano, 2008), which is consistent with findings in Rate (Siibak, 2009). Brand et al. (2012) studied profile images on dating sites, finding that image attractiveness significantly correlates to profile descriptions. In this previous study, females evaluated males' profile images and descriptions from the viewpoint of attractiveness.

Inspired by improvements in image processing techniques, researchers have attempted to extract computational characteristics from profile images and have used them to predict users' properties. Cristani et al. (2013) and Segalin et al. (2016) showed that features of user images labeled as "favourite" correlates with their personality traits, which are calculated in two ways: self-reported personality and rater-reported personality. Wei and Stillwell (2016) examined the relationship between Facebook profile image features and user's *measured intelligence (MI)* or *perceived intelligence (PI)*. Here, they extracted image features using Deformable Part Model (Kokkinos, 2012) and Face++ detector (Zhou et al., 2013). Also, *MI* is defined by the score from 20-item IQ test (Raven et al., 2003). *PI* is measured by human ratings of intelligence impression based on users' profile images. They showed that *PI* has stronger correlations than *MI* with the image features. Liu et al. (2016) inspected the relationship between characteristics of profile images on Twitter and the types of personality measured from Big Five Personality (McCrae and John, 1992). They gathered approximately 60,000 users and estimated personality scores based on users' tweets, and showed a lot of correlations between the features of profile images and the scores of the personality of the users. For example, users who use default profile images are less likely to be active. Redi et al. (2015) examined how the ambiance of bars or cafes on Foursquare is related to visitors' profile image. Ambiance is defined by 18 dimensions (e.g., "party", "social", or "creative"). The researchers extracted profile image features using Face++ detector (Zhou et al., 2013). Results show that the ambiance of "friendly" and "social" are the easiest to predict.

Weber and Mejova (2016) classified Twitter users into two groups of perceived appearance based on their profile images. They then investigated differences in Twitter use (e.g., number of followers, tweets) between the two groups. Tominaga and Hijikata (2015) investigated user activities related to following and posting actions per profile image categories.

Our present study aims to understand the relationship between profile images and user activities on Twitter. In contrast to the previous studies, we investigate user activities of retweeting, replying, attaching URL links, and using hashtags, reveal quantitative differences in the user activities across profile images categories. Three models are then built to predict profile image categories from user activities based on logistic regression, support vector machine (SVM), and random forest, noting the models of best performance.

3 Categories of profile images

3.1 Definition

We establish 13 categories of objects observed in users' profile images: "oneself", "self portrait", "hidden face", "associate", "different person", "letter", "logo", "otaku", "character", "animal", "object", "scene", and "default". Table 1 lists brief descriptions of all 13 categories. Moreover, we show sample profile images pertaining to these categories in Figure 1. In figures and tables appearing later in this paper, we use the codes shown in the center column of Table 1 to represent the listed categories (e.g., "On" stands for oneself). In the next subsections, we explain our user experiments to take these steps.
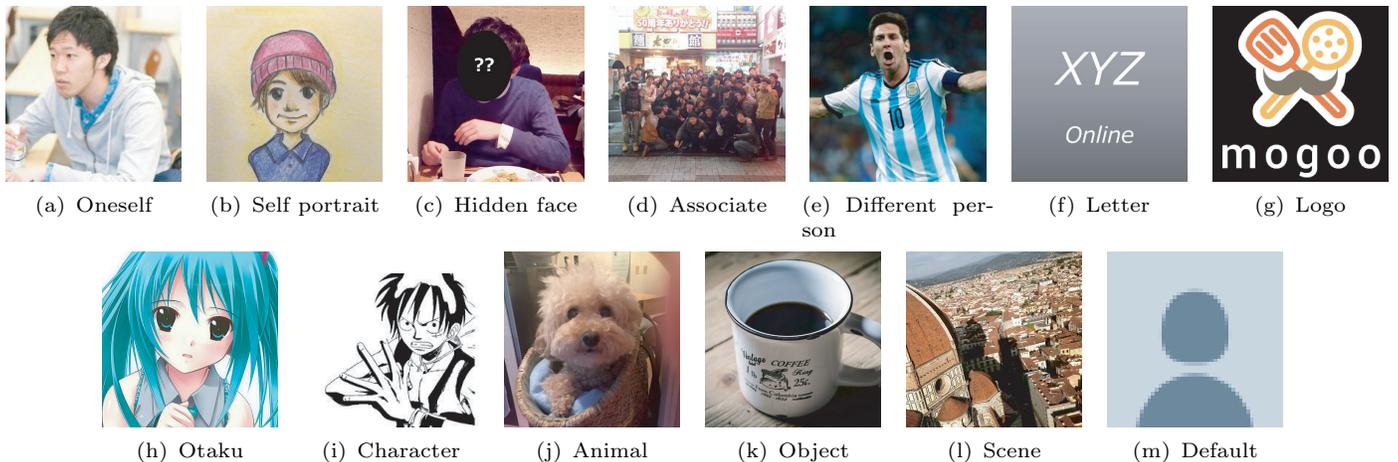
3.2 Validation of coincidence step

With our first experiment, we verify the coincidence of people's classification of profile images on Twitter. The procedure of this experiment is as follows.

First, we invited four coders: graduate students at Osaka University, receiving compensation for this experiment. Second, we used the Twitter Sample API to randomly gather Japanese Twitter users from September 18th, 2013 to October 17th, 2013. As a result, 20,833,001 tweets and 4,394,542 unique users were sampled. From this large pool of unique

Table 1: The categories, codes and descriptions of profile images

Category	Code	Description
Oneself	On	Picture of the user himself/herself
Self portrait	Sp	Illustration of the user's face
Hidden face	Hf	Picture of the user with some part of the face hidden
Associate	As	Picture of the user with other people (e.g., friends or family)
Different person	Dp	Picture of a person other than the user (e.g., a celebrity or a child)
Letter	Le	Image consisting only letters
Logo	Lo	Image of a logo
Otaku	Ot	Picture of beautiful female characters from Japanese anime or manga
Character	Ch	Picture of famous cartoon characters other than female characters from Japanese anime or manga
Animal	An	Picture of animals such as birds, cats, and dogs
Object	Ob	Picture of an object such as a ball, a bike, and a cup (usually the users' possessions)
Scene	Sc	Picture of a natural scenery
Default	De	Default image

(a) Oneself   (b) Self portrait   (c) Hidden face   (d) Associate   (e) Different person   (f) Letter   (g) Logo

(h) Otaku   (i) Character   (j) Animal   (k) Object   (l) Scene   (m) Default

Figure 1: Sample profile images of the 13 categories ((g) is the logo of *mogoo, Startouts Inc.*)

users, we randomly picked 300 users after excluding "out-of-service" users, whose accounts of these users are frozen by Twitter (therefore they cannot use Twitter and we cannot get any information about them). We then collected 300 profile images of the 300 users as a sample set for this experiment. During the period cited, users had an opportunity to change their profile images. However, Whitty et al. (2017) showed that only 10.8% of Twitter users changed their profile images once per month or more. Our data for this analysis were gathered for a period of one month; therefore, we assume that our results are not largely affected by the image-change phenomenon. Finally, the four coders were asked to classify the 300 profile images into the given categories.

After evaluating the coders' classifications, we obtain a Fleiss coefficient of 0.704, affirming that the coders' classifications substantially correlate.

3.3   Validation of coverage step

The second experiment verifies the coverage ability of the categories for general profile images on Twitter. Here, we investigate the extent to which profile images are included in any

categories. Before coding the profile images, we considered the statistical number of images needed, calculating a sample size of 1,067. The formula used is shown below.

$$n = \frac{N}{(\frac{c}{Z})^2 \cdot \frac{N-1}{p \cdot (1-p)} + 1} \tag{1}$$

In this calculation, we set the confidence level to 95%, giving us $Z = 1.96$; the confidence interval is 3.0, giving us $c = 0.03$; and the population N is 20,833,001, which is the number of tweets in our large pool. Moreover, we set $p = 0.50$, the most general setting.

As well as coincidence step, we firstly invited four coders: graduate students at Osaka University, receiving compensation for this experiment. We then asked each coder to categorize a different set of 300 profile images randomly selected from the large pool (section 3.2). We also created a category named "others", into which coders were to place profile images that do not belong to any other category. We found 113 out-of-service users in our sample and excluded them from this analysis. Finally, we obtained category labels for the 1,087 profile images, meeting the required sample size, $n \geq 1067$.

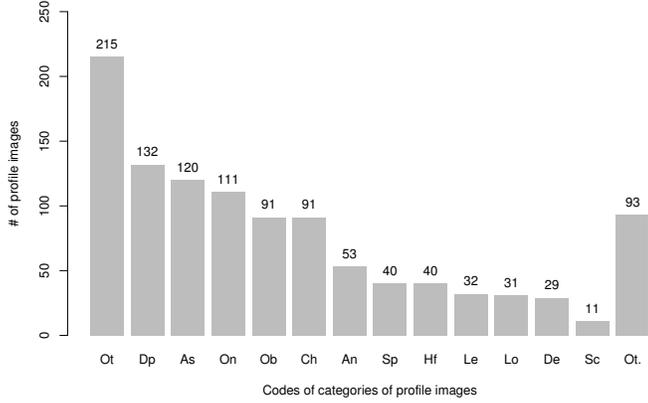Figure 2 shows the distribution of the 1,087 profile images

Figure 2: The distribution of the profile images classified by the four coders (Ot. means "others" category)

across the categories. Here, category names are represented in the horizontal axis by the codes shown in Table 1. The vertical axis represents the number of users whose profile image is categorized.

Of the 1,087 users, 93 are categorized into the "others" category ("Ot."). Here, the coverage ability is defined as the ratio of samples classified into one of our 13 categories to all 1,087 samples used. Accordingly, the coverage ability is 0.914. Therefore, the established 13 categories have a substantial ability to adequately classify general profile images on Twitter.

4 User activity

In this section, we explain the types of user activity in this study. When using Twitter, users may follow other users or post tweets. In this study, user action is expressed with the following functions: "follow," "post," "retweet," "reply," "attach URL links," and "attach hashtags ($\#$)." User activity is defined from user action. Specifically, user activity is represented as usage frequency or ratio of user actions.

FF **(ratio of followers to followees)**: Twitter users "follow" other users to collect information. Users generally start following others according to their interests or social relationships in the real world. Letting the number of followees be $followees$ and the number of followers be $followers$, we define a user activity FF as the ratio of $followers$ to $followees$.

$$FF = \frac{followers}{followees} \tag{2}$$

R_{tw} **(frequency of tweets per day)**: In addition to a "follow" action on Twitter, users "post" tweets to deliver information to others per their preferences. Letting the number of all tweets posted by a target user be $Tweets_{All}$ and the user lifetime be $UDays$, we define a user activity R_{tw} as the ratio of $Tweets_{All}$ to $UDays$, which means frequency of posting tweets a day.

$$R_{tw} = \frac{Tweets_{All}}{UDays} \tag{3}$$

R_{rt} **(frequency of retweets per tweets)**: If users wish to forward a followee's tweet to their followers, they use the "retweet" function if they believe the information is interesting or useful for followers. Moreover, as if the tweet was retweeted by many others, it would include valuable or interesting information for audience. We define two types of user activities related to a "retweet" action. First is R_{rt}, representing the ratio of the number of retweets to tweets during the data collection period. R_{rt} is calculated using $Retweets$ (the number of users' retweets) and $Tweets$ (the number of the users' tweets in the data collection period). Here, $Tweets$ is different from $Tweets_{All}$. Because of the Twitter REST API limitation, we cannot obtain more than 3,200 tweets from a single user. Thus, we separately acquire the number of tweets posted by the target user in the total lifetime and in the data collection period. Whereas the total number of a user's tweets is called $Tweet_{All}$, the number of the user's tweets within the data collection period is called $Tweets$. $Tweets_{All}$ is used for calculating R_{tw}, and $Tweets$ is used for the other user activities. R_{rt} is the frequency of retweets per tweet.

$$R_{rt} = \frac{Retweets}{Tweets} \tag{4}$$

R_{rted} **(frequency of retweeted tweets per tweets)**: The second one is R_{rted}, representing the ratio of the number of retweeted tweets to tweets in the data collection period: the frequency of retweets by others. As follows, R_{rted} is calculated using $Retweeted$, the number of retweeted tweets in the data collection period, and $Tweets$.

$$R_{rted} = \frac{Retweeted}{Tweets} \tag{5}$$

R_{reply} **(frequency of replies per tweets)**: If users create a conversation with other users on Twitter, they use the "reply" function. Therefore, the reply function is mainly used for making conversation. User activity, R_{reply}, is defined as the ratio of $Replies$ (the number of reply tweets in the data collection period) to $Tweets$, which describes the extent to which a target user prefers communicating with other users.

$$R_{reply} = \frac{Replies}{Tweets} \tag{6}$$

R_{hash} **(frequency of tweets with hashtags per tweets)**: If users wish to propagate their tweets to more users, they include "hashtags ($\#$)", which categorize tweets and make them easily searched and located by others. User activity R_{hash} is defined as the ratio of $Hashtags$ (the number of a user's tweets in the data collection period with hashtags) to $Tweets$, which represents frequency of using hashtags per tweet.

$$R_{hash} = \frac{Hashtags}{Tweets} \tag{7}$$

R_{url} **(frequency of tweets with url per tweets)**: Twitter has the 140-character limit, so, when users wish to deliver additional information, they attach URL links to their tweets. To

measure the use of this function, we define user activity, R_{url}, as the ratio of $URLs$ (the number of user tweets containing URL links from the data collection period) to $Tweets$, which provides the frequency of using URL links per tweet.

$$R_{url} = \frac{URLs}{Tweets} \qquad (8)$$

5   Method

5.1   Data collection

To quantify user activities, we require the data as follows: $followers$, $followees$, $Tweets_{All}$, $UDays$, $Tweets$, $Retweets$, $Retweeted$, $Replies$, $Hashtags$, and $URLs$. Data are collected as follows. First, we pick a target user randomly from the pool of four million users. Next, a target user's profile image is shown to the coders, who then classify it into one of 13 categories. The process is repeated until the number of users classified into each category reaches exactly 100. In case that the number of users classified into a specific category reaches 100, we asked the coders to discontinue classifying profile images into the category and discard the images. In total, the number of the target users reaches 1,300. Next, we acquire user data using the Twitter REST API from February 23, 2014 to March 25, 2014, calculating all user activities.

Table 2 summarizes medians and standard deviations of all user activities per the 13 categories. In total, standard deviations towards medians tend to be large because of outliers' influence. From the medians in Table 2, tendencies in user activities are roughly understood.

5.2   Analysis

First, we conduct a logistic regression analysis to understand which types of user activities are related to the profile image categories. Second, we build prediction models to determine whether users belong to a specific category based on their user activities. Both 2-way classifiers and 13-way classifiers are used. Finally, we estimate user activities from profile image selections, which is opposite to the method mentioned above. The detail of this procedure is explained below.

5.2.1   Effects of user activities on profile image categories

The purpose of this examination is qualitatively and statistically understanding how user activities correlate to profile image categories. Therein, we use logistic regression analysis to measure the effects of multiple independent variables on categorical dependent variables. In the context of our study, we examine the effects of user activities (e.g., FF or R_{tw}) on the categorical inclusion of profile images (i.e., true or false).

Regression analysis is conducted as follows. First, we divide all users into two groups, based on whether they are included in one target category. Next, all user activities are scaled so that the average is 0.0 and the deviation is 1.0. Finally, a logistic regression analysis is conducted for each profile image category.

5.2.2   Prediction of profile image categories from user activities

(A) 2-way classifier

Using logistic regression, random forest, and support vector machine (SVM), we build binary-classification models to predict profile images from user activities.

First, we select positive and negative datasets. When we choose a specific category as the target, we randomly pick up another 100 users from other categories. The reason behind the random user selection from other categories is avoiding *imbalanced-data problem* (Japkowicz, 2000). The 100 users and their user activities in the target category are defined as a positive dataset, and the another 100 users and their user activities are defined as a negative dataset.

Second, we divide the datasets into training and test sets. We then conduct a 10-cross validation to avoid over-fitting problems. Thus, the training set contains 90% of a positive and a negative dataset respectively. For the training dataset, we train the classification models. In the all machine learning techniques, the objective variable is positive or negative. Moreover, the predictor variables represent user activities. In each machine learning model, we tune for the best performance (e.g., selecting influential predictor variables or controlling internal parameters), as shown below.

Logistic regression: We select influential predictor variables based on *Akaike's Information Criteria (AIC)* (Akaike, 1973), which represents adaptability of built models. The variables selection aims to minimize *AIC*. Here, we adopt backward elimination method for the selection.

Random forest: This method measures *mean decrease Gini index* of each predictor variable. The Gini index reveals the extent of deviation of a classification result. The smaller the deviation, the better the classification result. Therefore, variables having larger *mean decrease Gini index* are regarded as better predictors. The number of variables for selecting the best performance is based on the *mean decrease Gini index*.

SVM: Instead of selecting influential variables, SVM tunes two internal parameters: *cost* and *gamma*, using grid search. *Cost* determines the extent of wrongly classified instances, and *gamma* represents boundary simplicity. An RBF Gaussian kernel is used for base conversion.

Finally, we evaluate the performance of each model using the test set with F-measure. This index is a harmonic mean of *precision* and *recall*. While *precision* means the ratio of actual target category users to those predicted, *recall* represents the ratio of predicted target category users to actual users belonging to the target category.

Noted that we randomly pick up 100 users as a negative dataset from the set of 1200 users in the other categories. The performance of the models might depend on the random selection. Therefore, we conduct the above process of preparing positive and negative datasets, training models, and evaluating the performance twelve times to select 100 users from 1200 users in the other categories.

Table 2: Medians and standard deviations of each user activity for all categories.

User activity	Oneself		Self portrait		Hidden face		Associate		Different person		Letter		Logo	
	Med.	S.D.	Med.	S.D.	Med.	S.D.	Med.	S.D.	Med.	S.D.	Med.	S.D.	Med.	S.D.
FF	1.137	1187.609	1.026	77.705	0.990	11881.358	0.994	3.926	0.964	1720.484	1.189	8748.564	2.000	1497.370
R_{tw}	3.483	15.043	4.550	21.747	9.067	29.614	7.867	8.414	5.650	22.367	2.783	21.475	2.200	13.991
R_{rt}	0.066	0.110	0.060	0.151	0.082	0.148	0.074	0.153	0.041	0.191	0.010	0.212	0.002	0.176
R_{rted}	0.049	0.248	0.102	0.268	0.021	0.153	0.042	0.605	0.030	0.219	0.210	0.346	0.316	0.345
R_{reply}	0.432	0.265	0.398	0.260	0.469	0.283	0.673	0.217	0.500	0.299	0.058	0.278	0.073	0.293
R_{url}	0.099	0.280	0.102	0.268	0.025	0.198	0.000	0.077	0.019	0.156	0.191	0.378	0.377	0.386
R_{hash}	0.000	0.141	0.015	0.159	0.018	0.150	0.000	0.024	0.002	0.073	0.000	0.224	0.004	0.263

User activity	Otaku		Character		Animal		Object		Scene		Default	
	Med.	S.D.	Med.	S.D.	Med.	S.D.	Med.	S.D.	Med.	S.D.	Med.	S.D.
FF	1.060	371.456	0.963	8.965	0.901	378.935	0.917	88.759	0.888	180.888	0.657	395.017
R_{tw}	26.033	36.187	5.200	20.734	2.317	22.860	6.167	28.211	2.583	19.062	0.300	22.865
R_{rt}	0.062	0.160	0.039	0.172	0.039	0.255	0.025	0.177	0.027	0.218	0.000	0.138
R_{rted}	0.026	7.366	0.019	1.338	0.000	0.195	0.007	0.141	0.001	0.227	0.000	0.208
R_{reply}	0.353	0.247	0.468	0.281	0.367	0.294	0.427	0.283	0.302	0.313	0.000	0.282
R_{url}	0.026	0.142	0.022	0.210	0.032	0.283	0.013	0.210	0.013	0.276	0.000	0.326
R_{hash}	0.032	5.245	0.005	0.154	0.000	0.199	0.000	0.089	0.000	0.151	0.000	0.203

(B) 13-way classifier

In addition to 2-way classifiers, we aim to build 13-way classifiers that predict the types of users' profile images based on user activities. For building prediction models, we use the same machine learning techniques as before.

Here, we adopt a 5-cross validation so that 20 users from each category are selected as a test set, because 13 classes are prepared and the number of users selected from each category should be larger than the number of classes. In addition to the 2-way classifier, we train classifiers using training sets as tuning parameters to the best performance. After training classifiers, we evaluate classifier performance. We show F-measures for all classes and a macro-averaging F-measure, which is a mean of F-measures for all classes, to examine overall classifier performance.

5.2.3　Estimation of user activities from profile image categories

The purpose of this analysis is to obtain insights about the extent to which we estimate user activities from users' profile images. To this end, we examine mean values and confidence intervals of user activities per categories of profile images. Results show the mean and an interval estimated by observed user activity data based on profile images. Therefore, we can estimate how users behave from their profile image selection. In this analysis, we use 95% confidence intervals.

6　Result and Implication

6.1　Effects of user activities on profile image categories

Table 3 shows the results of the logistic regression analysis. In this table, β is the partial regression coefficient, which describes polarity and magnitude of the effect of a specific independent variable on categorically dependent variables under the condition that other independent variables are constant.

In the category of "oneself," R_{tw} is found to provide a negative impact: ($\beta = -0.51$, $p < 0.01$). As discussed by Liu, et al., Twitter users whose profile images include only one face tend to be conscientious, preferring socially expected behaviors (Liu et al., 2016). Therefore, "oneself" users may use caution when posting to their Twitter communities. They also engage in replying activity ($\beta = 0.32$, $p < 0.01$) and URL link-attaching ($\beta = 0.35$, $p < 0.01$). As mentioned above, oneself users tend to behave as expected in social life (Liu et al., 2016); therefore, we can conclude that "oneself" users prefer maintaining social relationships by communicating with other users and by delivering detailed information.

On the contrary, "hidden face" users and "otaku" users are more likely to post tweets ($\beta = 0.31$, 0.71, $p < 0.001$ in both). Joinson stated that visual anonymity amplifies self-expression in computer-mediated communications when the people are strangers with each other (Joinson, 2001). This finding is consistent with the posting activities of "hidden face" users, because they hide their face in their profile images. Examining the median value of R_{tw}, shown in Table 2, we find that "otaku" users post at least one tweet per hour, on average. In Japan, tweets about *anime* (cartoon) or *manga* (comic book) during their respective TV broadcasts are called "live-broadcasting tweets." It is known that the number of tweets related to anime increases during its broadcast, and that the tweets are retweeted or replied-to not only during their broadcast but also afterwards (Aikawa et al., 2015). Manually checking, we find that these tweets tend to use fewer words than tweets from users in other categories. Here, we study the average tweet word count of users in the 13 categories. We exclude any usernames, hashtags, and URL links from users' tweets. We summarize the results in Table 4, showing medians of the average word count per the 13 categories. "Otaku" has the

Table 3: Results by logistic regression analysis showing the effect of user activities on categories of profile images (β is partial regression coefficient and S.E. is standard error. AIC means Akaike's Information Criteria (Akaike, 1973). $^{*}...p < 0.01$, $^{**}...p < 0.001$)

User activity	Oneself β	S.E.	Self portrait β	S.E.	Hidden face β	S.E.	Associate β	S.E.	Different person β	S.E.	Letter β	S.E.	Logo β	S.E.
FF	-0.17	0.31	-1.75	2.11	0.18	0.09	-123.09	118.65	0.00	0.12	0.15	0.07	-0.12	0.18
R_{tw}	**-0.51** *	0.18	-0.13	0.12	**0.31** **	0.08	-0.43	0.18	-0.04	0.11	-0.07	0.12	-0.43	0.17
R_{rt}	-0.26	0.13	-0.06	0.12	0.03	0.11	0.09	0.13	0.04	0.11	**0.56** **	0.13	0.15	0.14
R_{rted}	-0.32	0.60	-0.02	0.15	-3.03	1.55	0.02	0.12	-0.30	0.66	-0.02	0.12	0.01	0.10
R_{reply}	**0.32** *	0.12	0.12	0.12	0.16	0.12	**1.00** **	0.13	0.27	0.11	**-0.91** **	0.18	**-0.43** *	0.16
R_{url}	**0.35** *	0.11	0.19	0.11	-0.22	0.17	**-1.81** *	0.56	-0.45	0.19	**0.36** **	0.09	**0.63** **	0.09
R_{hash}	-1.97	1.19	-0.25	0.65	-0.06	0.28	-14.52	6.69	-1.67	1.57	-0.29	0.54	-0.02	0.14
AIC	682.13		712.53		670.41		555.84		695.03		644.68		627.18	

User activity	Otaku β	S.E.	Character β	S.E.	Animal β	S.E.	Object β	S.E.	Scene β	S.E.	Default β	S.E.
FF	-0.16	0.53	-44.06	37.55	-0.24	0.68	-0.92	2.73	-0.94	1.56	-0.11	0.33
R_{tw}	**0.71** **	0.08	-0.08	0.11	-0.22	0.13	0.18	0.09	-0.24	0.13	-0.03	0.11
R_{rt}	0.04	0.14	-0.14	0.13	0.26	0.10	0.00	0.12	0.26	0.11	-0.13	0.21
R_{rted}	0.58	0.27	0.05	0.06	**-3.60** *	1.36	-4.21	1.72	-1.84	1.14	**-4.71** *	1.45
R_{reply}	-0.03	0.14	0.25	0.11	-0.15	0.13	-0.02	0.12	-0.30	0.13	**-0.94** **	0.16
R_{url}	**-0.87** **	0.22	-0.13	0.14	0.15	0.12	-0.05	0.15	0.06	0.12	0.02	0.11
R_{hash}	**3.50** **	0.76	-0.21	0.66	-0.10	0.31	-1.78	1.44	-1.12	1.04	-0.67	0.84
AIC	570.08		696.27		697.23		695.60		699.50		631.03	

smallest median: 24.70. Thus, we conclude that "otaku" users post many live-broadcasting tweets using a small number of words. Considering this, it is intuitive that they are less likely to use URL links to convey detailed information ($\beta = -0.87$, $p < 0.001$).

In the category of "associate", R_{reply} wield a significantly positive role in $\beta = 1.00$ and $p < 0.001$. This result means that they are motivated to communicate with more users, or to communicate more often with fewer users. Because profile images of "associate" users include other faces, such as friends or family, this motive might influence their usage styles and profile image choices.

Whereas the effect of R_{url} on "oneself" category is significantly positive, as shown above, the effect on "associate" category is significantly negative: $\beta = -1.81$ and $p < 0.01$. The role of URL links in tweets is providing more detailed information. Users of "oneself" and "associate" categories reflect different frequencies of detailed information delivery to others. However, users in both categories are found to prefer communicating with other users via the "reply" function. One reason for both the difference in using URL links and the similarity in replying between these categories is their differing primary target audiences. There is a possibility that "oneself" users communicate with users with weak ties (e.g., business colleagues), and that "associate" users mainly interact with users with strong ties (e.g., friends and school mates). This speculation is supported by a previous study (Liu et al., 2016), which shows that the number of faces in profile pictures are negatively associated with "openness to experience," one of personality traits in the five factor model (McCrae and Costa, 1987). In other words, compared to "associate" users, "oneself" users interact with more new friends on Twitter. Also, it is shown that "extraversion" is positively correlated with the number of faces in profile pictures (Liu et al., 2016), meaning that "oneself" users are not so active to use Twitter. Considering these results, we

conclude that the users with weaker ties need more detailed information to feebly maintain their relationships. However, we cannot assess the validity of our speculation at this stage. So, we plan to correlate target audiences to categories of profile images in a future study.

Regarding "letter" and "logo" categories, R_{url} has a significantly positive effect: $\beta = 0.36$ and 0.63; $p < 0.001$ in both. Moreover, we find that R_{reply} negatively correlates to "letter" ($\beta = -0.91$, $p < 0.001$) and "logo" categories ($\beta = 0.43$, $p < 0.01$). Those categories contain many company and group accounts. Therefore, they may aim to advertise themselves on Twitter using URL links. Moreover, it is understood that accounts of companies or social groups are likely to avoid interacting with specific individuals, because they might not value promoting individual relationships.

In addition to "posting" activity, "otaku" users actively use hashtags ($\beta = 3.50$, $p < 0.001$). As discussed, "otaku" users are likely to post tweets about their preferences. Thus, they may attach hashtags to their tweets to share with others having similar interests, but are not yet friends.

According to Liu et al. (2016), default users (i.e., those who use default profile images) are inclined to reflect lower "extraversion" than any other personality traits in the five factor model (McCrae and Costa, 1987). Additionally, it is possible that those users are novice users, and are not yet active. Almost all user activities cause negative impacts with respect to this category. Particularly, R_{rted} and R_{reply} are significant ($\beta = -4.71, -0.94$, $p < 0.01, 0.001$). Generally, users do not perform these actions if they are not connected to many others. The median number of followees and followers of "default" users is relatively low (39.5 and 21.0, respectively).

At this stage, we cannot infer the cause of R_{rted} having a negative effect on the "animal" category ($\beta = -3.60$, $p < 0.01$). To understand it, our future work should investigate their motives for selecting the profile images.

Table 4: Medians of average word count in tweets per the 13 categories

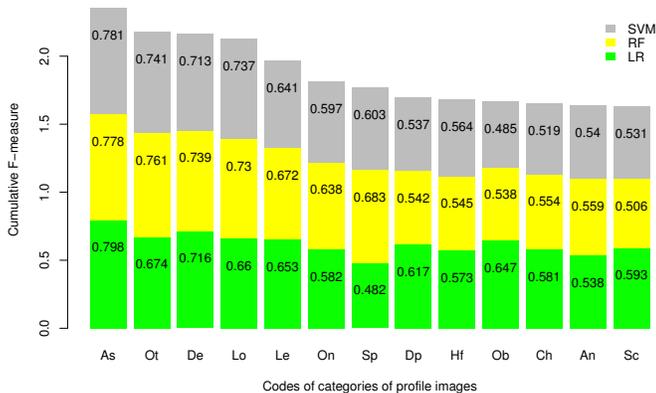| Code of categories	On	Sp	Hf	As	Dp	Le	Lo	Ot	Ch	An	Ob	Sc	De
Median	38.96	38.47	26.89	25.71	27.94	56.31	61.49	24.70	32.19	34.52	31.54	39.66	34.81

Figure 3: Cumulative F-measures by logistic regression, random forest, and SVM for all categories of profile images (2-way classifiers). LR and RF are logistic regression and random forest respectively.

6.2 Prediction of profile image categories from user activities

6.2.1 2-way classifiers

Figure 3 shows the F-measures of each classification model for all categories of profile images. Categories are shown in descending order of the cumulative F-measure. For simplicity, we describe the F-measure of a machine learning model, m, for a category, c, as f_c^m. Here, m is defined as lr, rf, or svm, which respectively stand for logistic regression, random forest, or SVM. c is defined by a category code shown in Table 1 of section 3.1. For example, the F-measure of logistic regression model for the "oneself" category is represented as $f_{On}^{lr} = 0.582$.

In all machine learning models, categories ranked in the top four are the same: "associate," "otaku," "default," and "logo." The "associate" category shows the best F-measures of all the models ($f_{As}^{lr} = 0.798$, $f_{As}^{rf} = 0.778$, $f_{As}^{svm} = 0.781$). In the logistic regression models for "associate" category, R_{reply} and R_{url} are selected as predictor variables during all ten validations, and these partial regression coefficients result in significantly positive and negative correlations, respectively, which is consistent with the results shown in Table 3. Moreover, with respect to the random forest models for "associate" category, R_{reply} is selected as the most influential variable when the algorithm builds the model. We cannot inspect the effect of each predictor variable on the performance of the SVM models; however, it is inferred that R_{reply} or R_{url} contributes to improving classification ability when we consider the results from the logistic regression and random forest models. It can also be said that R_{reply} and R_{url} reflect the performance of the classification model for the "associate" category, because those users are more likely to engage in replying and are less likely to add URL links to their tweets.

Other than the "associate" category, "otaku" ($f_{Ot}^{lr} = 0.674$, $f_{As}^{rf} = 0.761$, $f_{As}^{svm} = 0.741$), "default" ($f_{De}^{lr} = 0.716$, $f_{De}^{rf} = 0.739$, $f_{De}^{svm} = 0.713$), and "logo" ($f_{Lo}^{lr} = 0.660$, $f_{Lo}^{rf} = 0.730$, $f_{Lo}^{svm} = 0.737$) categories also show high F-measures. Constantly selected variables are significantly effective predictors for categories in logistic regression models. The variables for "otaku" are R_{tw}, R_{url}, and R_{hash}. Moreover, whereas R_{rted} and R_{reply} are elected for "default," R_{reply} and R_{url} are elected for "logo" and "associate." These results are consistent with Table 3. Concerning the random forest models for these categories, the selected variables are R_{tw} for "otaku," R_{FF} for "default," and R_{url} for "logo." Results shown in Table 2 are useful for better understanding the selected variables in the random forest for these categories. Values of the selected variables for "otaku," "default," and "logo" are considerably higher or lower than those from the other categories. Generally, random forest hierarchically classifies given instances into two categories per a threshold value of each feature. Therefore, R_{tw}, R_{FF} and R_{url} provide a valid classification threshold for "otaku," "default," and "logo," respectfully.

In contrast to the categories with the top four performances, "hidden face," "character," "animal," and "scene" do not show any superior performances in any machine learning model ($f_{Hf} < 0.573$, $f_{Ch} < 0.581$, $f_{An} < 0.559$, $f_{Sc} < 0.593$). Results shown in Table 3 indicate that "character" and "scene" do not reflect any significantly effective user activities. Therefore, the models cannot identify unique features of these categories, leading to worse performance. In Table 3, "hidden face" and "animal" are respectively unique in R_{tw} and R_{rted}. However, "otaku" and "default" also show significant coefficients for the same user activities. Moreover, "hidden face" and "animal" users perform less unique activities than "otaku" and "default" users. Therefore, "hidden face" and "animal" categories might pose difficulties for the machine learning models when distinguishing "hidden face" users from "otaku" users and "animal" users from "default" users.

By comparing the overall performance of each machine learning model, we find that random forest is the best, SVM is second best, and logistic regression is third best. Performance differences among the machine learning models stem from performance tuning strategies. When building models, logistic regression aims to minimize AIC, which is defined by the summation of model adaptability and the number of selected predictor variables. Here, model adaptability is a residual sum of squares between the ground truth and prediction. If the algorithm cannot find effective variables, model adaptability is not sufficiently decreased by selecting variables. Therefore, the algorithm selects a few variables to minimize AIC, because it does not sufficiently improve adaptability. Thus, we conclude that logistic regression is likely to produce the worst performance from this dataset.

Random forest and SVM are better than logistic regression. As described in Liu et al. (2013), random forest, proposed by
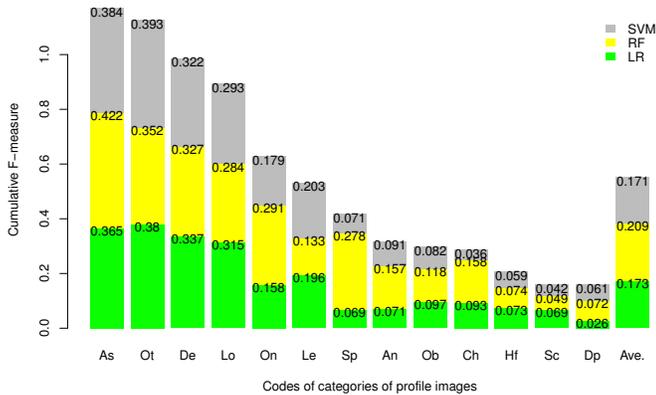
Figure 4: Cumulative F-measures by logistic regression, random forest, and SVM for all categories of profile images (13-way classification). "Ave." is the mean of F-measures across all categories, per the machine learning models.

Breiman (2001), directly improves classification quality with models using the *Gini index*. Therefore, it is relatively easy for random forest to achieve higher performance for classification tasks. SVM, proposed based on the statistical learning theory by Cortes and Vapnik (1995), maps a space of given features to a higher-dimensional space with base conversion. It then draws a discrimination plane, balancing boundary simplicity of the plane, defined as γ, with the proportion of wrongly-classified instances, defined as C. Tuning these parameters, SVM strives to achieve better performance by generating an abstract space from given features. Thus, it captures the overall combination of unique variables. Accordingly, the performance achieved by SVM models is likely to be higher and more stable.

Regarding the F-measure results from random forest and SVM, categories ranked in the upper half and lower half are the same. However, the difference in F-measures between upper and lower halves of random forest is larger than in SVM. As mentioned above, random forest directly improves the classification quality. Therefore, if it cannot seize distinctive variables, the performance inevitably becomes worse. Alternatively, from top to bottom, F-measures smoothly decrease in SVM models, regardless of the upper and lower halves. Because of the learning strategy of SVM, its performance tends to be stable, and the upper half of SVM model F-measures does not considerably differ from the lower half.

6.2.2 13-way classifiers

Figure 4 shows cumulative F-measures of each category by 13-way classifiers per the machine learning models. In Figure 4, a cumulative bar located at the right, named "Ave.," shows the mean of F-measures across all categories per the machine learning models. This mean value is called *macro-Fscore* (Sokolova and Lapalme, 2009), which represents overall performance of multi-class classifiers. The baseline is a random classifier and its F-measure is 1/13 (0.077), because the number of classes is 13 for each classifier and we prepared the same number of users in each category (both of precision and recall of random classifiers are 1/13). Similar to the 2-way classifiers, we show

the F-measures calculated by the 13-way classifiers as $mul\text{-}f_c^m$, where m and c are mean types of machine learning models and categories of profile images, respectively.

"Associate" (As), "otaku" (Ot), "default" (De), and "logo" (Lo) categories show better performance. Among all models of all categories, the random forest model for "associate" shows the best performance ($mul\text{-}f_{As}^{rf}$=0.422). As mentioned before, these categories reflect unique user activities and combinations. Thus, 13-way classifiers also easily identify users per their categories.

Concerning to overall performance of all models, random forest is the best ($mul\text{-}f_{\bar{c}}^{rf} = 0.209$), as well as for the 2-way classifiers. As mentioned in 6.2.1, this machine learning technique directly improves classification results based on the *Gini index*. Thus, it can consistently show better performance over the other machine learning techniques, plus, it tackles multi-classification problems.

Compared to the 2-way classifiers, differences in performance among categories are relatively larger. If a category does not possess many unique user activities, classifiers cannot obtain clues to know whether users belong to a given category. Thus, users in a category that does not perform many unique activities are likely to be predicted as users from other categories. For example, the number of users predicted as belonging to target categories is sometimes 0 for "hidden face," "scene," or "different person."

In contrast to the two-way classifiers, "otaku" shows better performance than "associate" using logistic regression and SVM models ($mul\text{-}f_{Ot}^{lr}$, $mul\text{-}f_{As}^{lr} = 0.380, 0.365$, $mul\text{-}f_{Ot}^{svm}$, $mul\text{-}f_{As}^{svm} = 0.393, 0.384$). Examining the prediction results in detail, we find that the number of predicted users in "associate" tend to be larger in these two machine learning models than in the random forest models. This causes model precision to be low. Therefore, we find lower F-measures in "associate." At this stage, we cannot provide insights into the reason why precision is worse. However, we will tackle this issue in future work.

6.3 Estimation of user activities from profile image categories

In this section, we discuss the extent to which we can estimate user activities from categories of target user profile images. To this end, we calculate means and confidential intervals (CIs) of user activities per the 13 categories. Figure 5 shows the mean values (plots) and 95% CI (dotted lines) of target user activities in each category, and aligns the 13 categories from small (left) to large (right), in terms of CI range. The y-axis fits the range of mean values; confidence intervals are partly cut off.

Results demonstrate that the CI of user activities for categories tend to reflect larger ranges if categories show higher mean values in these activities. In other words, it is difficult to estimate frequent user activities with high probability. For example, in Figure 5(a), "oneself" (On), "different person" (Dp), "letter" (Le), or "hidden face" (Hf) categories have higher mean values and wider CI ranges of FF. Compared to these four categories, "logo" (Lo) has a smaller CI range and a higher mean value. Thus, predicting FF of "logo" users ought to be easier. For "associate" (As), "character" (Ch), or
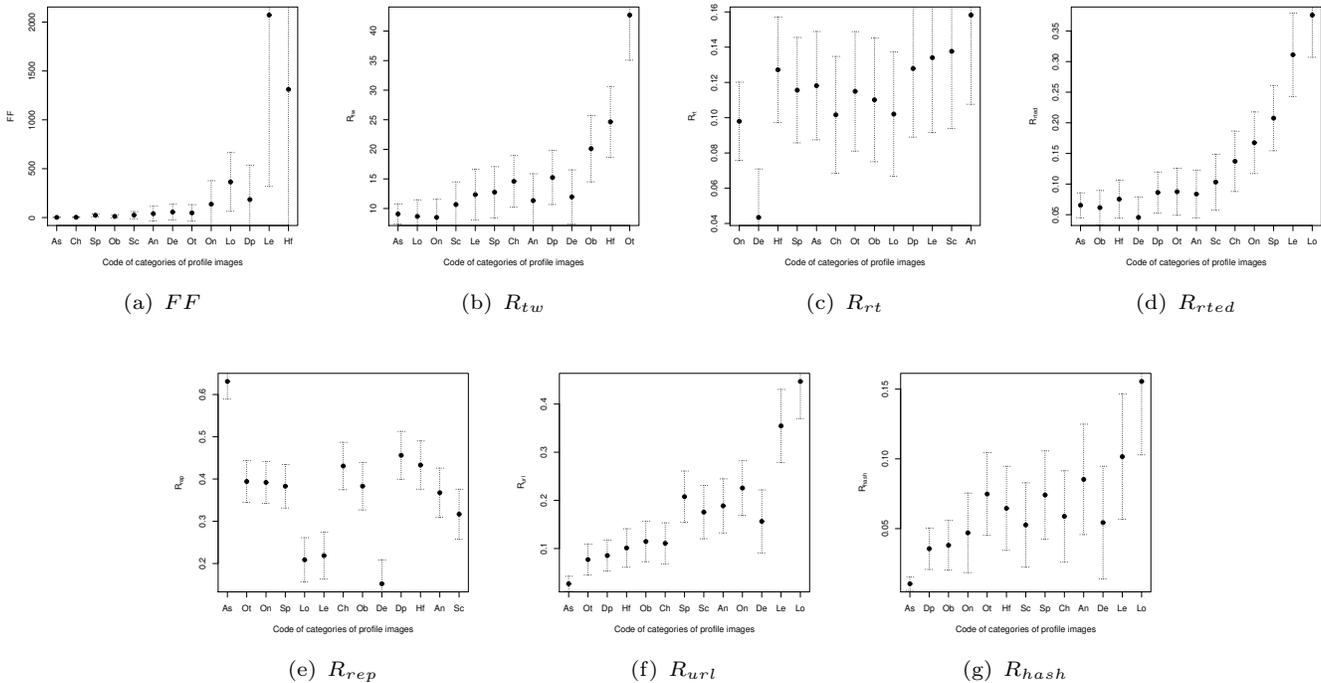
Figure 5: Means and 95% confidence intervals of each user activity per the 13 categories. Plots and lines represent mean values and confidence intervals.

"self portrait" (Sp) categories, mean values of FF are lower, which correspond to lower CI ranges. Twitter users in these three categories tend to balance their follower-to-followee ratio. Otherwise, these users are likely to engage in reciprocal relationships, following someone after the person follows them, or vice versa. Possibly, we predict FF of users in these categories with relatively high probability.

Additionally, as shown in Figure 5(b), 5(d), 5(f), and 5(g), R_{tw}, R_{rted}, R_{url}, and R_{hash} also demonstrate that larger mean values are likely to reflect wider CI ranges. For these activities, "associate" (As) category shows the smallest CI ranges. As with FF, we might estimate user activities of "associate" users with high probability. "Otaku" (Ot) shows the highest mean value and the widest CI range of R_{tw}, the daily frequency of posting tweets (42.70 ± 7.63, Figure 5(b)). Compared to other categories, highly accurate prediction for posting frequency of "otaku" users is relatively difficult. However, on average, we estimate that "otaku" users post tweets once or twice per hour. Concerning R_{rted}, R_{url}, and R_{hash}, estimating the frequency of user activities for "logo" (Lo) and "letter" (Le) users ought to be easier than FF, because their CI ranges are smaller.

We also find that there are no large variances in CI ranges of R_{rt} and R_{rep} across the categories, as shown in Figure 5(c) and 5(e). Interestingly, "associate" (As) category shows the highest mean value and the smallest CI range of R_{rep} (0.63 ± 0.04). This result shows that most "associate" Twitter users massively engage in replying activities. Therefore, if we detect that a new user chooses a group photo as a profile image, the user might often reply to other users. Specifically, the estimated frequency of replying is approximately five-to-

seven times per day, which is calculated from statistics shown in Figure 5(b) and 5(e).

Our results can be used to tackle cold-start problem (Schein et al., 2002) in Twitter user recommendations, because when a new user sets a profile image, we can estimate user activity based on the given profile image. For example, if a new user is found to select an "associate" photo as a profile image, the user would be comfortable if users who prefer communication with the associate user such as real-world friends or family members were recommended. New "logo" or "letter" users will frequently post tweets with URL links or hashtags, and will reply two-to-four times per day. Therefore, it might be useful for new "logo" or "letter" users to receive recommendations to connect with similar users; they would learn faster how to promote their content from the behaviors of the recommended users. Moreover, they do not prefer in-person communication on Twitter. Therefore their reply messages to someone should not be promoted so that these messages do not automatically appear in timelines of non-follower users.

7   Limitation

Here we discuss limitations of our study design to generalize our findings.

First, all target users in this study are Japanese. It is reported that usage motives and social media patterns differ by cultural background (Kim et al., 2011; Vasalou et al., 2010). Therefore, it is possible that the tendency to choose profile images depends on users' cultural peculiarities. We are now

conducting a cross-cultural study to categorize types of profile images. This study will help us generalize our findings.

Second, we only deal with basic user attributes, like, number of followers, tweets, and so forth. However, we do not inspect contents of the posted tweets. By inspecting tweets, we may be able to measure more discernable differences. For instance, "logo" and "letter" users may narrow their tweets to specific topics. However, "associate" users may mention various topics in their conversation.

Moreover, our results may be influenced by target user demographics. For example, users who show themselves in their profile images are more likely to be older, inferring a high likelihood of posting URLs. There may be more females in the "associate" category, which might correlate to frequent replies. In our future work, we will inspect the effects of users' demographic information.

Finally, we find that reply-frequency is an important predictor of profile image categories. However, we have not yet inspected who replies to the target users. Considering the characteristics of replying partners, we might improve classification ability of each machine learning model. For instance, it is likely that "oneself" users frequently communicate with business colleagues, but that "otaku" users frequently interact with cyber-world friends. In our future work, we address this phenomenon.

8 Conclusion

In this study, we examined how Japanese Twitter user activities correlate to categories of profile images. First, we sorted the users into 13 categories based on the nature of their profile images: "oneself," "self-portrait," "hidden face," "associate," "different person," "letter," "logo," "otaku," "character," "animal," "object," "scene," and "default." Then, per the categories, we investigated seven types of user activities related to "following" or "followed," "tweeting," "retweeting," "retweeted," "replying," "tweeting with URL links," and "tweeting with hashtags." Finally, we statistically analyzed how user activities differ by the categories. Furthermore, we built machine learning models to predict categories of profile images from user activities using logistic regression, random forest, and SVM.

We found that several categories of profile images significantly correlate to user activities. For example, "otaku" users prefer posting tweets frequently, "associate" users are likely to communicate with other users, "logo" users tend to add URL links to their tweets, and "default" users are less likely to reply to other users. Alternatively, categories such as "character" or "scene" do not reflect unique user activities. All machine learning models perform best for "associate" ($f_{As}^{lr} = 0.798$, $f_{As}^{rf} = 0.778$, $f_{As}^{svm} = 0781$). For "default," "otaku," and "logo" categories, the models also show good performance ($f_{De}^{lr} = 0.716$, $f_{De}^{rf} = 0.739$, $f_{De}^{svm} = 0713$; $f_{Ot}^{lr} = 0.674$, $f_{Ot}^{rf} = 0.761$, $f_{Ot}^{svm} = 0741$; $f_{Lo}^{lr} = 0.660$, $f_{Lo}^{rf} = 0.730$, $f_{Lo}^{svm} = 0737$). These results indicate that it is easy for the models to predict categories in which users display unique activities. Comparing the overall performance of the machine learning techniques, we find that SVM is the best.

In addition to 2-way classifiers, we find that identifying "associate" users is the easiest for the 13-way classifiers. However, compared to 2-way classifiers, there is a larger difference in overall performance among the 13 categories. We examined prediction of user activities from profile images and find that user activities are predicted more easily if a user selects a profile image from an "associate" category. We also discussed the usability of the results from this examination.

Several shortcomings remain in this study design, such as the fact that only Japanese users were sampled, and that there was no examination of the tweet contents. To overcome these limitations, we plan to conduct a cross-cultural study to examine cultural differences in Twitter profile images, and to conduct a content analysis to understand the types of topics tweeted. We hope that our examination and analysis will provide more insights and will be useful to other studies. We also believe that this study is the first step in understanding the relationship between profile images and user activities.

References

Abel, F., Q. Gao, G.-J. Houben, and K. Tao (2011a), 'Analyzing user modeling on Twitter for personalized news recommendations'. *User Modeling, Adaption and Personalization* pp. 1–12.

Abel, F., Q. Gao, G. J. Houben, and K. Tao (2011b), 'Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web'. *The Semantic Web: Research and Applications* pp. 375–389.

Aikawa, M., K. Senkai, C. Yasuda, H. Sakai, Y. Ogawa, and E. Umehara (2015), 'Analysis of Tweets of Midnight Anime Programs through Twitter'. *Forum on Information Technology* **14**(4), 85–92.

Akaike, H. (1973), 'Information theory and an extension of the maximum likelihood principle'. In: *Proceedings of 2nd International Symposium on Information Theory*. pp. 267–281, Akademiai Kiado.

Brand, R. J., A. Bonatsos, R. D'Orazio, and H. Deshong (2012), 'What is beautiful is good, even online: Correlations between photo attractiveness and text attractiveness in men's online dating profiles'. *Computers in Human Behavior* **28**(1), 166–170.

Breiman, L. (2001), 'Random Forests'. *Machine learning* **45**(1), 5–32.

Cortes, C. and V. Vapnik (1995), 'Support-Vector Networks'. *Machine Learning* **20**(3), 273–297.

Cristani, M., A. Vinciarelli, C. Segalin, and A. Perina (2013), 'Unveiling the Multimedia Unconscious: Implicit Cognitive Processes and Multimedia Content Analysis'. In: *Proceedings of the 21st ACM International Conference on Multimedia*. pp. 213–222, ACM.

Fong, K. and R. A. Mar (2015), 'What Does My Avatar Say About Me? Inferring Personality From Avatars'. *Personality and Social Psychology Bulletin* **41**(2), 237–249.

Goffman, E. (2002), 'The presentation of self in everyday life. 1959'. *Garden City, NY*.

Hannon, J., M. Bennett, and B. Smyth (2010), 'Recommending Twitter Users to Follow Using Content and Collaborative Filtering Approaches'. In: *Proceedings of the Fourth ACM Conference on Recommender Systems*. New York, NY, USA, pp. 199–206, ACM.

Hum, N. J., P. E. Chamberlin, B. L. Hambright, A. C. Portwood, A. C. Schat, and J. L. Bevan (2011), 'A picture is worth a thousand words: A content analysis of Facebook profile photographs'. *Computers in Human Behavior* **27**(5), 1828–1833.

Ikeda, K., G. Hattori, C. Ono, H. Asoh, and T. Higashino (2013), 'Twitter user profiling based on text and community mining for market analysis'. *Knowledge-Based Systems* **51**, 35–47.

Japkowicz, N. (2000), 'Learning from Imbalanced Data Dets: A Comparison of Various Strategies'. In: *AAAI Workshop on Learning from Imbalanced Data Sets (AAAI'00)*. pp. 10–15.

Joinson, A. N. (2001), 'Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity'. *European Journal of Social Psychology* **31**(2), 177–192.

Kapidzic, S. (2013), 'Narcissism as a predictor of motivations behind Facebook profile picture selection.'. *Cyberpsychology, Behavior and Social Networking* **16**(1), 14–19.

Kim, Y., D. Sohn, and S. M. Choi (2011), 'Cultural difference in motivations for using social network sites: A comparative study of American and Korean college students'. *Computers in Human Behavior* **27**(1), 365–372.

Kokkinos, I. (2012), 'Bounding Part Scores for Rapid Detection with Deformable Part Models'. In: *Proceedings of the 12th International Conference on Computer Vision - Volume Part III*. pp. 41–50, Springer-Verlag.

Kwak, H., C. Lee, H. Park, and S. Moon (2010), 'What is Twitter, a Social Network or a News Media?'. In: *Proceedings of the 19th International Conference on World Wide Web*. New York, NY, USA, pp. 591–600, ACM.

Liu, L., D. Preotiuc-Pietro, Z. R. Samani, M. E. Moghaddam, and L. Ungar (2016), 'Analyzing Personality through Social Media Profile Picture Choice'. In: *Proceedings of the Tenth International AAAI Conference on Web and Social Media*.

Liu, M., M. Wang, J. Wang, and D. Li (2013), 'Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar'. *Sensors and Actuators B: Chemical* **177**, 970–980.

McCrae, R. R. and P. T. Costa (1987), 'Validation of the Five-Factor Model of Personality Across Instruments and Observers'. *Journal of Personality and Social Psychology* **52**(1), 81–90.

McCrae, R. R. and O. P. John (1992), 'An introduction to the five-factor model and its applications.'. *Journal of personality* **60**(2), 175–215.

Pan, Y., F. Cong, K. Chen, and Y. Yu (2013), 'Diffusion-aware Personalized Social Update Recommendation'. In: *Proceedings of the 7th ACM Conference on Recommender Systems*. New York, NY, USA, pp. 69–76, ACM.

Ramage, D., S. Dumais, and D. Liebling (2010), 'Characterizing Microblogs with Topic Models'. In: *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. pp. 130–137.

Raven, J., J. C. Raven, and J. Court (2003), 'Manual for Raven's progressive matrices and vocabulary scales'. *San Antonia: Harcourt Assessment*.

Redi, M., D. Quercia, L. T. Graham, and S. D. Gosling (2015), 'Like Partying? Your Face Says It All. Predicting the Ambiance of Places with Profile Pictures'. In: *Proceedings of the 9th International AAAI Conference on Web and Social Media*. pp. 347–356.

Sakaki, T., M. Okazaki, and Y. Matsuo (2010), 'Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors'. In: *Proceedings of the 19th International Conference on World Wide Web*. New York, NY, USA, pp. 851–860, ACM.

Schein, A. I., A. Popescul, L. H. Ungar, and D. M. Pennock (2002), 'Methods and metrics for cold-start recommendations'. In: *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '02*. New York, New York, USA, pp. 253–260, ACM Press.

Segalin, C., A. Perina, M. Cristani, and A. Vinciarelli (2016), 'The Pictures we Like are our Image: Continuous Mapping of Favorite Pictures into Self-Assessed and Attributed Personality Traits'. *IEEE Transactions on Affective Computing* **PP**(99), 1–1.

Siibak, A. (2009), 'Constructing the Self through the Photo Selection - Visual Impression Management on Social Networking Websites'. *Cyberpsychology: Journal of Psychological Research on Cyberspace* **3**(1).

Siibak, A. (2010), 'Constructing masculinity on a social networking site'. *Young* **18**(4), 403–425.

Siibak, A. and P. Hernwall (2011), ''Looking like my favourite Barbie' – Online Gender Construc- tion of Tween Girls in Estonia and in Sweden'. *Studies of Transition States and Societies* **3**(2), 57–68.

Sokolova, M. and G. Lapalme (2009), 'A systematic analysis of performance measures for classification tasks'. *Information Processing and Management* **45**(4), 427–437.

Steele Jr, F., D. C. Evans, and R. K. Green (2009), 'Is your profile picture worth 1000 words? Photo characteristics associated with personality impression agreement.'. In: *Third International AAAI Conference on Weblogs and Social Media.* pp. 327–330.

Strano, M. M. (2008), 'User Descriptions and Interpretations of Self-Presentation through Facebook Profile Images'. *Cyberpsychology: Journal of Psychological Research on Cyberspace* **2**(2).

Tominaga, T. and Y. Hijikata (2015), 'Study on the Relationship Between Profile Images and User Behaviors on Twitter'. In: *Proceedings of the 24th International Conference on World Wide Web.* New York, NY, USA, pp. 825–828, ACM.

Turner, M. and N. Hunt (2014), 'What Does Your Profile Picture Say About You? The Accuracy of Thin-Slice Personality Judgments from Social Networking Sites Made at Zero-Acquaintance'. In: *Proceedings of the 6th International Conference on Social Computing and Social Media - Volume 8531.* pp. 506–516, Springer.

Vasalou, A., A. N. Joinson, and D. Courvoisier (2010), 'Cultural differences, experience with social networks and the nature of "true commitment" in Facebook'. *International Journal of Human Computer Studies* **68**(10), 719–728.

Wang, S. S., S. I. Moon, K. H. Kwon, C. A. Evans, and M. A. Stefanone (2010), 'Face off: Implications of visual cues on initiating friendship on Facebook'. *Computers in Human Behavior* **26**(2), 226–234.

Weber, I. and Y. Mejova (2016), 'Crowdsourcing Health Labels: Inferring Body Weight from Profile Pictures'. In: *Proceedings of the 6th International Conference on Digital Health Conference.* pp. 105–109, ACM.

Wei, X. and D. Stillwell (2016), 'How smart does your profile image look? Estimating intelligence from social network profile images'. In: *arXiv 1606.09264.*

Weng, J. and B.-S. Lee (2011), 'Event Detection in Twitter'. In: *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.* pp. 401–408.

Whitty, M. T., J. Doodson, S. Creese, and D. Hodges (2017), 'A picture tells a thousand words: What Facebook and Twitter images convey about our personality'. *Personality and Individual Differences.* doi:10.1016/j.paid.2016.12.050.

Zhao, C. and G. Jiang (2011), 'Cultural Differences on Visual Self-presentation Through Social Networking Site Profile Images'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* New York, NY, USA, pp. 1129–1132, ACM.

Zhao, W. X., J. Jiang, J. Weng, J. He, E.-p. Lim, H. Yan, and X. Li (2011), 'Comparing Twitter and traditional media using topic models'. *Advances in Information Retrieval* pp. 338–349.

Zhou, E., H. Fan, Z. Cao, Y. Jiang, and Q. Yin (2013), 'Extensive facial landmark localization with coarse-to-fine convolutional network cascade'. In: *Proceedings of the IEEE International Conference on Computer Vision.* pp. 386–391.

Tomu Tominaga graduated the master course of Graduate School of Engineering Science at Osaka University in 2016. He is currently a Ph.D. student in Graduate School of Engineering Science at Osaka University. He was a visiting scholar in GroupLens Research at University of Minnesota in 2017. He engages in HCI research including social computing, computational social science, and computer-supported cooperative work.

Yoshinori Hijikata graduated the master course of Osaka University in engineering science in 1998. After graduation, he joined IBM research, Tokyo research Laboratory. He received Ph.D. in 2002 from Osaka University. He was assistant professor (2002-2006) and associate professor (2007-2017) in Osaka University. He was a visiting researcher in GroupLens Research of University of Minnesota in 2014. Currently, he is associate professor in Kwansei Gakuin University. His research interests are social computing including recommender systems and social media analysis. He takes a unique approach by combining a computer science and social science for examining the users' behaviors and social events.