

Photos Don't Have Me, But How Do You Know Me?

Analyzing and Predicting Users on Instagram

Kyungsik Han
Ajou University
Republic of Korea
kyungsikhan@ajou.ac.kr

Yonggeol Jo
Ajou University
Republic of Korea
yonggeol93@ajou.ac.kr

Youngseung Jeon
Ajou University
Republic of Korea
jeonyoungs@ajou.ac.kr

Bogoan Kim
Ajou University
Republic of Korea
bokim1122@ajou.ac.kr

Junho Song
Hanyang University
Republic of Korea
junonjuno@agape.hanyang.ac.kr

Sang-Wook Kim
Hanyang University
Republic of Korea
wook@agape.hanyang.ac.kr

ABSTRACT

This paper presents a comprehensive analysis on social media use and engagement by age and gender on Instagram. We define five user age groups (from 10s to 50s) and two user gender groups (males and females), and compare them based on three aspects: *activity*, *image object*, and *tag*. We especially excluded the information that indicates human (e.g., selfies, faces, body) for each aspect in order to examine whether users are still identifiable without that information. Our study results indicate that each user group exhibits unique characteristics and the features from each aspect can be used to develop user classification models (82% for gender and 43% for age classification) without relying on the information that specifically indicates age and gender.

CCS CONCEPTS

• **Human-centered computing** → **User models; Social media; Web-based interaction; Hypertext/hypermedia**; • **Social and professional topics** → **User characteristics**;

KEYWORDS

User modeling, age and gender, social media, Instagram, comparison analysis

ACM Reference Format:

Kyungsik Han, Yonggeol Jo, Youngseung Jeon, Bogoan Kim, Junho Song, and Sang-Wook Kim. 2018. Photos Don't Have Me, But How Do You Know Me?: Analyzing and Predicting Users on Instagram. In *UMAP'18 Adjunct: 26th Conference on User Modeling, Adaptation and Personalization Adjunct, July 8–11, 2018, Singapore, Singapore*. ACM, New York, NY, USA, Article 4, 6 pages. <https://doi.org/10.1145/3213586.3225232>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UMAP'18 Adjunct, July 8–11, 2018, Singapore, Singapore

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-5784-5/18/07...\$15.00

<https://doi.org/10.1145/3213586.3225232>

1 INTRODUCTION

Social media has been penetrated into people's daily life, allowing people to access, create, and interact with a wide range of information [15, 17]. This aspect can be observed across users with different ages and genders, because social media provides users with an easy-way of creating and maintaining friendships and sharing their own stories or interests to their friends or the public in many formats, including text messages, images, or videos and in diverse social media platforms [9]. This is also accelerated by smartphone adoption, allowing people to be engaged in social media at anytime and in anywhere, greatly satisfying their needs [19]. This all leads them to spending a significant amount of time on using social media. Teens and young adults are among the earliest social media adopters and continue to use these sites at high levels, and usage by older adults has increased in recent years. Statistically in case of Facebook, 88% of people aged between 18 and 29, 84% of people between 30-49, and 72% of people older than 50 use Facebook [7].

A great body of research has characterized age- and gender-based user groups from the perspectives of use and engagement. Using an ethnographic approach, Quinn et al. [24] found that younger users (ages 15-30, including teens) tend to use different social media features (e.g., updating their status, etc.) more often than older users (ages over 50). Dhir et al. [5] provided empirical evidence on how adolescents (ages 12-19), young adults (ages 20-30), and adults (ages 31-50) differ in terms of selfie behavior by studying 3,763 users. Their results showed that, compared to adults, adolescents are more likely to take individual and group selfies, post their own selfies, and use image-graphic filters. They also found that young adults are more likely than older adults to take individual and group selfies, and post and edit images. Pfeil et al. [21] found that teens tend to express themselves and share their stories in online spaces more than older users (ages over 60). Teens also tend to use multiple social networking sites and maintain multiple channels of communication more often than older users because each channel offers different methods of interacting with others. Han et al. [11] identified teens (ages 13-19) and adults (ages 30-39) and examined their behavioral differences with respect to the number of images posted, number of Likes and comments received, popularity of the user, and so on. Using these features, they demonstrated that two user groups could be identified with up to 81% accuracy. Regarding

gender differences, many studies have examined the linguistic characteristics. For example, Argamon et al. [1] and Rao et al. [25] found that females tend to use more words related to social interaction while males use more formal, affirmation, and informational words.

Although prior research, including the ones discussed above, has identified various characteristics of user groups in social media, research has somewhat limited to relying on (1) two types of the age groups, limiting the age diversity [10, 11, 14, 21], (2) specific activities or posting behaviors such as posting selfies [5], or (3) text-based information [1, 25]. Research lacks the comprehensive analysis on various behavioral characteristics (e.g., image content, tagging, commenting, liking) of more diverse user groups as well as differences in behaviors among those user groups.

To gain additional insights on the relationship between user groups and their social media behaviors, this paper explores the characteristics of the users from various age and gender groups. In this paper, we chose Instagram for data collection, because it is one of the most popular social media platforms.

We report usage differences in Instagram among five age groups (from 10s to 50s) and two gender groups (male and female), and present the implications of the differences in social media. Especially, the data we analyzed and used for modeling *do not contain the information that specifies a user's age and gender*. For example, selfies, human-related objects or tags were all excluded. With these data, we strive to answer the following research questions.

- What are the activity-, tag-, and image-based characteristics of user age groups based on their usage patterns on Instagram? How are those characteristics different across the groups?
- Can we build a classification model that predicts a user's age and gender without relying on the information that specifies age and gender information?

Overall, our study results show (1) clear differences in activities, image objects, and tags among the age and gender groups, and (2) a fairly high classification performance of our proposed model. This paper contributes to offering not only data-driven, straightforward insights on the image-sharing behaviors of users in social media, but also the potential for identifying specific user age or gender groups. Our findings can be applied to identifying users who have similar interests or engage in similar activities, offering personalized recommendations, and detecting abnormal activities that will help users engage with social media in a more interactive and meaningful fashion.

2 DATA COLLECTION

Figure 1 illustrates the overall procedure of our study. The first phase was data collection. We chose Instagram, because it is one of the most popular social networking sites and has high levels of user engagement. According to a recent Pew Research report in 2018, 35% of all U.S. adults use Instagram [27], an increase of seven percentage points from the 28% who said they did in 2016 [8]. Because of its high popularity, there have been a substantial number of studies on Instagram. Some example studies are exploring the relationships between image content and engagement [2], analyzing image content and user types [12], studying tag-based Like

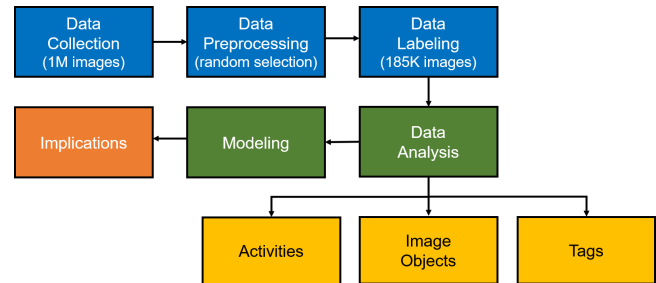


Figure 1: The procedure of the study. After data collection and manual labeling, we analyzed the characteristics of the age and the gender user groups based on activities, image objects, and tags.

Table 1: Summary of the data for age group.

Age group	Users	Images	Image objects	Tags
10s	527	26,350	25,442	97,562
20s	809	40,450	38,981	156,630
30s	789	39,450	37,822	195,177
40s	977	48,850	46,618	249,063
50s	604	30,200	28,246	139,431
Total	3,706	185,300	177,203	837,863

Table 2: Summary of the data for gender group.

Age group	Users	Images	Image objects	Tags
Male	753	37,650	35,476	168,630
Female	2,953	147,650	141,727	669,233
Total	3,706	185,300	177,203	837,863

networks formed by users who share common tags [13], focusing on selfie categories using tags and clustering [4], etc.

Our method of data collection is using tags. We noted that some of the tags imply user demographics: age or gender. For this reason, we used a list of tags that directly indicate poster's age information such as #im15, #iam32yearsold, #my49thbirthday, etc. We developed a web crawler using Python, which was designed to retrieve the images that contain the tag that we want to search for and other meta-data, including the number of Likes and comments added, posted time, image URL, etc. In addition, we collected user profile information, including user profile image, bio description, the number of images posted, the number of followers, etc. As a result, we initially collected approximately 1,000,000 images.

3 EXPERIMENT PROCEDURE

3.1 Age and gender labeling

After data collection, we manually annotated the age and gender information for each user by looking at their profile information and posted images. Three authors of the paper carefully looked into the profile and posted images of the users and filtered out the ones if there is no consensus between the three authors. As a result, Tables 1 and 2 summarize the data collected and used in the

Table 3: Results of comparative analyses for different user groups. (* $p < 0.05$).

	Age - Median (Mean)						Gender - Median (Mean)		
	10s	20s	30s	40s	50s	η^2	M	F	η^2
Like count	29.0 (65.8)	26.0 (50.0)	22.0 (57.2)	20.0 (36.1)	19.0 (30.8)	0.01*	25.0 (46.6)	22.0 (47.1)	0.00
Tag count	1.03 (2.01)	1.02 (1.70)	1.05 (2.36)	1.03 (1.92)	1.02 (1.83)	0.01*	1.00 (2.02)	1.00 (1.97)	0.00
Comment count	1.02 (3.41)	2.00 (3.66)	3.08 (4.95)	3.07 (5.11)	3.05 (4.71)	0.02*	2.00 (4.48)	3.00 (4.45)	0.00
Caption length	31.0 (60.7)	36.0 (65.9)	44.0 (108.4)	39.0 (77.5)	39.0 (79.4)	0.02*	36.0 (68.5)	38.0 (82.3)	0.00
Emoji count	0.02 (1.21)	0.01 (1.11)	0.00 (1.02)	0.00 (0.73)	0.03 (0.71)	0.01*	0.00 (0.58)	0.00 (1.05)	0.01*
Emoji ratio	0.01 (0.08)	0.00 (0.07)	0.00 (0.04)	0.00 (0.03)	0.00 (0.03)	0.02*	0.00 (0.02)	0.00 (0.06)	0.01*
Emoji presence	0.00 (0.50)	0.00 (0.50)	0.00 (0.37)	0.00 (0.26)	0.00 (0.26)	0.05*	0.00 (0.22)	0.00 (0.41)	0.03*
Emoji unique	0.01 (1.00)	0.00 (0.91)	0.00 (0.84)	0.00 (0.54)	0.00 (0.53)	0.02*	0.00 (0.44)	0.00 (0.83)	0.02*

analysis for age and gender, respectively. We have a total of 3,706 user samples, 185,300 images, 177,203 image objects, and 837,863 tags.

3.2 Image object retrieval

To retrieve image information, we used the Google Cloud Vision Application¹ Programming Interface (API). It uses machine learning models to classify images into thousands of categories, detect individual objects and faces within images, and find and read printed words contained within the images. It offers several detection features, including label detection, face detection (showing four types of emotions, including joy, sadness, anger, and surprise), landmark detection, and logo detection. We used label detection for the analysis. Each identified object in the image comes with a confidence level (i.e., accuracy), ranging from 0.0 to 1.0 (higher level means greater confidence). We used image objects with accuracies greater than 70%.

3.3 Excluding human-related features

Our goal is to examine the difference in characterizing and modeling user groups by excluding human related information. There are two reasons for this. First, detecting a user’s age and gender information with the images, that explicitly include his/her faces and people in general, would be too obvious. Second, many users on Instagram (and broadly social media) share images without including their faces or information that directly indicate their age or gender. Hence, excluding human-related information for the characterization of users is expected to give interesting insights that have not yet been explored and understood.

To do that, we removed all tags and image objects that indicate a user’s age and gender information. The total number of removed content is 304, and some examples include selfie, smile, forehead, girl, boy, man, people, person, hair, nose, chin, forehead, me, child, portrait, and so on. The tags used for data collection (e.g., #im15yearsold) were also removed.

3.4 Activity features

We define the types and justification of activity-based features as follows.

- ‘Like Count,’ ‘Tag Count,’ ‘Comment Count,’ and ‘Caption Length’: These are some of the selected features to explain the key elements of the posts on Instagram.
- ‘Emoji’: We have considered emojis as one of the features, given that they are widely used by people and likely to offer the characteristics of different generations and both genders [20]. Additional traits of emojis that we analyzed include ‘Emoji count,’ the percentage of emojis used in the overall caption (‘Emoji ratio’), whether or not emojis have been used in the text (‘Emoji presence’), and the variety of unique types of emojis (‘Emoji unique’).

4 RESULTS

In this section, we report the results of individual aspect (activity-, image-, and tag-based analysis) and of classification models.

4.1 Activity-based analysis

Table 3 shows the summary of Instagram activities for each group. Because all variables show a long-tailed distribution, we considered both the median and the mean values for the analysis. Due to the large and non-normalized datasets, to correctly measure the differences, we used eta-square (the effect size), denoted as η^2 , which refers to the proportion of variance associated with each of the main effects, interactions, and error in an ANOVA study [29]. As a rule of thumb, $\eta^2 = 0.01, 0.13, 0.26$ are considered to be small, medium and large respectively.

The results present several interesting insights.

For age differences, we can see that the number of Likes gradually decreases from 10s to 50s, while that of comments generally increases. It appears that the Like interactions are active among the younger age groups (e.g., 10s and 20s), while the comment interactions are more active among the older age groups (e.g., 30s, 40s, and 50s). In addition, the length of the captions tends to be shorter in the 10s and 20s than in the 30s, 40s, and 50s. The differences in the four emoji-related features are noticeable among the user groups. From 10s to 50s, all features gradually decrease, indicating the high usage of emojis in the younger age groups.

Regarding gender differences, we did not find meaningful results except the four emoji-features. Apparently, females tend to post more emojis than males, implying that their expressions are more diverse and emotional.

¹<https://cloud.google.com/vision/>

Table 4: Top 10 image objects by the age and gender groups. For age, a blue cell indicates the objects shown more in the 10s and 20s groups, and a pink cell indicates the objects shown more in the 30s, 40s, and 50s. The two gender groups showed similar results, while a yellow cell highlights the objects relate to males, and a purple cell highlights objects relate to females.

Age - Object (%)					Gender - Object (%)	
10s	20s	30s	40s	50s	M	F
product(3.29)	fun(3.38)	fun(2.96)	fun(2.86)	fun(2.53)	fun(2.56)	fun(3.08)
fun(3.07)	product(2.82)	product(2.82)	text(2.81)	text(2.17)	product(2.41)	product(2.78)
text(2.33)	photography(2.17)	text(2.51)	product(2.63)	sky(2.17)	sky(2.21)	text(2.37)
photography(2.33)	sky(1.72)	font(2.24)	font(2.45)	product(2.03)	text(2.04)	font(2.05)
font(1.93)	text(1.59)	sky(1.56)	sky(1.84)	font(1.89)	font(1.76)	photography(1.67)
snapshot(1.47)	font(1.35)	photography(1.41)	tree(1.39)	tree(1.53)	tree(1.47)	sky(1.61)
sky(1.36)	snapshot(1.24)	day(1.37)	photography(1.19)	photography(1.14)	photography(1.37)	day(1.21)
day(1.25)	tree(1.23)	tree(1.09)	day(1.12)	plant(1.12)	water(1.12)	tree(1.21)
purple(1.19)	day(1.05)	pink(1.01)	food(1.01)	food(1.11)	vehicle(1.04)	pink(1.02)
photograph(1.06)	photograph(1.03)	food(0.98)	water(1.0)	water(1.08)	car(0.98)	snapshot(1.02)

Table 5: Top 10 user-posted tags by the age and gender groups. A yellow cell refers to a tag related to ‘emotion,’ ‘expression,’ a pink cell refers to a tag related to ‘fitness,’ ‘wellbeing,’ a green cell refers to a tag related to ‘nature,’ and a blue cell refers to ‘instagram-specific tags.’

Age - Tag (%)					Gender - Tag (%)	
10s	20s	30s	40s	50s	M	F
love(2.22)	love(2.49)	love(3.22)	love(2.75)	love(2.74)	love(1.46)	love(2.93)
cute(1.66)	happy(1.14)	fitness(1.34)	nofilter(0.97)	fitness(1.23)	tbt(1.19)	happy(0.99)
like4like(1.23)	cute(1.01)	happy(1.02)	beautiful(0.8)	summer(0.98)	art(1.04)	beautiful(0.87)
followme(1.22)	beautiful(0.99)	goals(0.92)	happy(0.74)	tbt(0.96)	summer(0.93)	summer(0.76)
happy(1.21)	summer(0.93)	fitfam(0.88)	repost(0.73)	catsofinstagram(0.89)	instagood(0.91)	fitness(0.74)
follow(1.01)	tbt(0.78)	blessed(0.87)	fitness(0.72)	nofilter(0.85)	fitness(0.87)	cute(0.75)
beautiful(0.99)	makeup(0.68)	travel(0.79)	tbt(0.68)	nature(0.84)	beautiful(0.77)	tbt(0.68)
likeforlike(0.99)	art(0.63)	beautiful(0.77)	instagood(0.64)	travel(0.78)	happy(0.76)	nofilter(0.62)
instagood(0.86)	photography(0.57)	beachbody(0.75)	summer(0.63)	beach(0.69)	nature(0.69)	blessed(0.56)
summer(0.82)	instagood(0.56)	tbt(0.71)	fun(0.58)	sunset(0.67)	nofilter(0.66)	travel(0.55)

4.2 Image object-based analysis

Table 4 shows a summary of top 10 image objects that appear in the images by user group. First, regardless of age and gender, all user groups have several same image objects, including product, fun, text, and font. Second, when we look at the age groups, a snapshot object only appears in the 10s and the 20s groups, while the objects related to nature (i.e., sky, tree, plant) show more in the 40s and the 50s groups. Third, regarding the gender groups, the male group has vehicle and car objects, while the female group has the pink and snapshot objects.

4.3 Tag-based analysis

Similar to the image-based analysis (Table 4), Table 5 shows a summary of the top 10 tags that appear in the images by each user group. First, the tags related to emotions or expressions (e.g., love, cute, happy, blessed) are presented over all age and gender groups. Second, regarding the age group, tags that are popularly used on Instagram (i.e., instagram specific tags; likedlike, followme, follow, instagood, throw-back-thursday (tbt), etc.) more appear in the 10s group. Tags related to fitness or wellbeing (e.g., fitness, fitfam, beachbody) are more presented by the 30s group, and those related to nature (e.g., summer, nature, sunset) more appear in the 50s group. Third, regarding the gender group, the male and the

female groups generally present a similar set of the tags, yet some specific tags related to emotions or expressions (e.g., love, happy, beautiful, cute) are the top ranked tags for the female group.

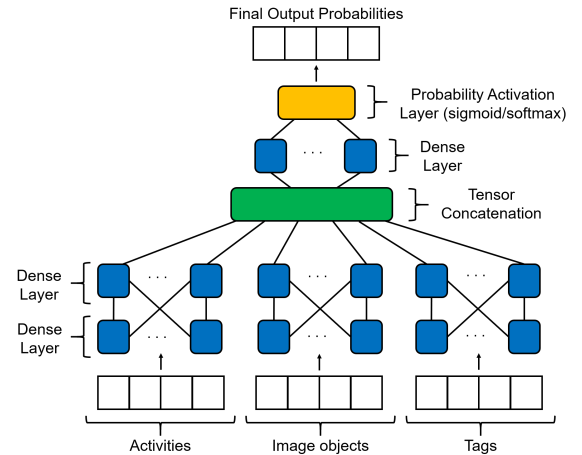


Figure 2: Neural network architecture for classification (MLP). The model combines three separate sub models built based on the features of activities, image objects, and tags.

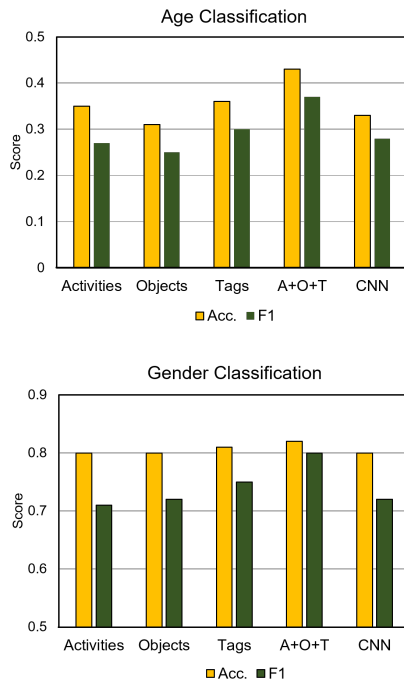


Figure 3: Performance of the classification model (MLP) for gender and age excluding any human-related features. The MLP that combines three individuals models yielded the best performance for both gender and age classification, and better performance than the CNN model.

4.4 Classification Model

Our goal of classification is to examine whether we can build classification models based on the types of activities that we defined, image objects, and/or tags. We trained age and gender-group classification models using a dataset that consists of the vectors of activities, image objects and tags.

We used term frequency inverse document frequency (TF-IDF) [26] and created the vector space with various feature sizes (100, 500, and 1000). Since we found that the models with 500 features yielded the greatest performance, we report the results with the 500 vectors.

For machine learning models, we used classic algorithms (i.e., logistic regression (LR) and support vector machine (SVM)) as well as neural network algorithms (i.e., multi-layer perceptron (MLP), convolutional neural network (CNN)), which are widely-used machine learning algorithms that yield overall good performance. Here we report the best results from various parameter choices.

For building a neural network model (Figure 2), we relied on the late fusion approach, that has been shown to be effective in many machine learning tasks [16]. "Fusion" allows for a network to learn a combined representation of multiple input streams. In our case, this fusion was executed in the classification layers. To combine these networks into one prediction and train them together, we merged these dense layers before the final classification. We trained our models for 100 epochs using the ADAM optimization algorithm [18]. We used 5-fold cross validation (i.e., 80% for training and 20% for testing, repeated 5 times) to predict age and gender. We used the

keras library², one of the widely used open source deep learning libraries, for modeling.

The performance results of the classification models are presented in Figure 3. For both age and gender detection, the merged model yielded the best performance (age: 43% (accuracy), 37% (F1 score); gender: 82% (accuracy), 80% (F1 score)). The other models with activities, objects, or tags exhibited similar or slightly greater performance with the one with CNN as a baseline.

5 DISCUSSION

Data that represents human engagement on social media provides ample opportunities to understand how humans behave online. Using this rationale, we aimed not only to understand the behavioral patterns of user groups through the comparative analysis of large datasets of activities, image objects, and tags, but also to create prediction models for classifying different age and gender groups. In addition, we presented an in-depth, comparative analysis of user image-sharing practices using deep learning techniques.

One of our study results indicate clear differences among the age groups, especially between the younger user groups (10s/20s) and the older user groups (40s/50s). Prior research has reported that the young generations present different motivations and expectations when using social media. The high Likes, emoji activities, and the top used tags and identified image objects represent the social and emotional aspects of the young age groups. For instance, boyd [3] presents that teens consider social media to be a space to model their identity and status and maintain cultural cues by interacting with the content posted by their friends or people they are following or themselves. Pfeil et al. [22] report that users in their 10s and 20s tend to express themselves and share own stories in online space more than older users. Young generations use diverse social media platforms and maintain multiple forms of communications because each channel provides different ways of interacting with others. They also quickly switch to and adopt other platforms if they provide better and unique features [23]. On the other hand, the older age group tends to describe their posts in more detail and leave more comments. In addition, their posts tend to have more professionally-taken images (e.g., nature, travel, atmosphere), related to their specific hobbies, well-being, etc. For gender differences, although less obviously shown compared to age, females tend to present more emotional content (e.g., happy, beautiful, many emojis) than males. This somewhat reflected stereotypical results and indicated that the gender differences inherent in many behaviors are mirrored in online spaces [28].

These behavioral characteristics of social media use, which are based on user properties, allowed us to build classification models that were found to achieve fairly good performance. The models can be further tuned and improved, yet our main objective of this paper is to demonstrate a potential of detecting a user's age and gender without having the information that specifically indicates age and gender. We found that (the 10s and the 20s groups) and (the 40s and the 50s groups) exhibited similar patterns. Hence, classification models that define the younger and older age groups possibly yield greater performance. Our next step is to build models that consider various characteristics and behavioral patterns of users.

²<https://keras.io/>

The followings are the limitations of the present study, which will be addressed in the future study. First, our study results may not be generalized, because the original user samples used for classification indicate their age information to some extent (either on their profile or through tags). Our next step is to collect more data for both age and gender user groups and build more comprehensive analyses and classification models. We will identify users and collect data that represent diverse types of users on social media. Collecting a greater number of samples is necessary as well. We will also include other activity types, such as temporal information (posted time, weekday and weekend), topics from image objects and tags, etc.

Second, another future work includes performing more comprehensive analysis by combining the results of image processing and text (e.g., captions, tags, comments) to extract more precise content from images by using additional services for image recognition (e.g., MS Azure Cognitive Service, Amazon Rekognition) that include other properties of images based on their own models. The analysis could include a study on temporal changes in image sharing activities, identification of image objects that potentially lead to privacy leakage, relationships with personality [6], etc.

6 CONCLUSION

Since the behavior patterns that users present in social media are very diverse, it is necessary to analyze the users based on various perspectives. As the use of image-based social media explosively increases, image-based user information, that is less obviously presented in the text-based content, is being generated and can be used to understand the user from a new point of view. Our study results highlight the possibility of identifying a user's basic trait – age and gender – with reasonably high accuracies, without relying on the information that specifically indicates age and gender. Many people use social media, while disclosing themselves. At the same time, there are people who use social media without revealing themselves, thus understanding various social media users is important in that this gives researchers and practitioners a broad understanding of social media use and engagement. This not only leads to understanding societal interests (e.g., activities in the online space and network connections), but also being utilized as various application services (e.g., providing social media users with customized services tailored to them, recommending other users with similar age, gender, and interests). We hope that our research will be one step in helping to achieve this.

ACKNOWLEDGMENTS

This work was supported by the new faculty research fund of the Ajou University [S-2017-G0001-00040]; National Research Foundation of Korea (NRF) grant [NRF-2017R1A2B3004581].

REFERENCES

- [1] Shlomo Argamon, Moshe Koppel, Jonathan Fine, and Anat Rachel Shimoni. 2003. Gender, Genre, and Writing Style in Formal Written Texts. *TEXT-THE HAGUE THEN AMSTERDAM THEN BERLIN* 23, 3 (2003), 321–346.
- [2] Saeideh Bakhshi, David A Shamma, and Eric Gilbert. 2014. Faces Engage Us: Photos With Faces Attract More Likes And Comments on Instagram. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 965–974.
- [3] Danah Michele boyd. 2008. *Taken out of Context: American Teen Sociality in Networked Publics*. University of California, Berkeley.
- [4] J. Deeb-Swihart, C. Polack, E. Gilbert, and I. A. Essa. 2017. Selfie-Presentation in Everyday Life: A Large-Scale Characterization of Selfie Contexts on Instagram. In *Proceedings of the International Conference on Web and Social Media*. 42–51.
- [5] Amandeep Dhir, Ståle Pallesen, Torbjørn Torsheim, and Cecilie Schou Andreassen. 2016. Do Age and Gender Differences Exist in Selfie-Related Behaviours? *Computers in Human Behavior* 63 (2016), 549–555.
- [6] Bruce Ferwerda, Markus Schedl, and Marko Tkalcić. 2015. Predicting Personality Traits with Instagram Pictures. In *Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems 2015*. ACM, 7–10.
- [7] Shannon Greenwood, Andrew Perrin, and Maeve Duggan. 2016. Social Media Update 2016. *Pew Research Center* 11 (2016).
- [8] Shannon Greenwood, Andrew Perrin, and Maeve Duggan. 2016. Social Media Update 2016. (2016).
- [9] Rebecca E Grinter, Leysia Palen, and Margery Eldridge. 2006. Chatting with Teenagers: Considering the Place of Chat Technologies in Teen Life. *ACM Transactions on Computer-Human Interaction (TOCHI)* 13, 4 (2006), 423–447.
- [10] Rita G Guimaraes, Renata Lopes Rosa, Denise De Gaetano, Demosthenes Z Rodriguez, and Graca Bressan. 2017. Age Groups Classification in Social Network Using Deep Learning. *IEEE Access* (2017).
- [11] Kyungsik Han, Sanghack Lee, Jin Yea Jang, Yong Jung, and Dongwon Lee. 2016. Teens Are From Mars, Adults Are From Venus: Analyzing and Predicting Age Groups With Behavioral Characteristics in Instagram. In *Proceedings of the 8th ACM Conference on Web Science*. ACM, 35–44.
- [12] Yuheng Hu, Lydia Manikonda, Subbarao Kambhampati, et al. 2014. What We Instagram: A First Analysis of Instagram Photo Content and User Types. In *Proceedings of the International Conference on Web and Social Media*.
- [13] Jin Yea Jang, Kyungsik Han, and Dongwon Lee. 2015. No Reciprocity in Liking Photos: Analyzing Like Activities in Instagram. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. ACM, 273–282.
- [14] Jin Yea Jang, Kyungsik Han, Patrick C Shih, and Dongwon Lee. 2015. Generation Like: Comparative Characteristics in Instagram. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 4039–4042.
- [15] Andreas M Kaplan and Michael Haenlein. 2010. Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons* 53, 1 (2010), 59–68.
- [16] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. 2014. Large-Scale Video Classification with Convolutional Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1725–1732.
- [17] Hyoungkoo Khang, Eyun-Jung Ki, and Lan Ye. 2012. Social Media Research in Advertising, Communication, Marketing, and Public Relations, 1997–2010. *Journalism & Mass Communication Quarterly* 89, 2 (2012), 279–298.
- [18] Diederik P Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [19] Mary Madden, Amanda Lenhart, Sandra Cortesi, Urs Gasser, Maeve Duggan, Aaron Smith, and Meredith Beaton. 2013. Teens, Social Media, and Privacy. *Pew Research Center* 21 (2013), 2–86.
- [20] Petra Kralj Novak, Jasmina Smaliović, Borut Sluban, and Igor Mozetič. 2015. Sentiment of Emojis. *PloS One* 10, 12 (2015), e0144296.
- [21] Ulrike Pfeil, Raj Arjan, and Panayiotis Zaphiris. 2009. Age Differences in Online Social Networking—A Study of User Profiles and the Social Capital Divide Among Teenagers and Older Users in MySpace. *Computers in Human Behavior* 25, 3 (2009), 643–654.
- [22] Ulrike Pfeil, Raj Arjan, and Panayiotis Zaphiris. 2009. Age Differences in Online Social Networking—A Study of User Profiles and the Social Capital Divide Among Teenagers and Older Users in MySpace. *Computers in Human Behavior* 25, 3 (2009), 643–654.
- [23] Anabel Quan-Haase and Alyson L Young. 2010. Uses and Gratifications of Social Media: A Comparison of Facebook and Instant Messaging. *Bulletin of Science, Technology & Society* 30, 5 (2010), 350–361.
- [24] Darren Quinn, Liming Chen, and Maurice Mulvenna. 2011. Does Age Make a Difference in the Behaviour of Online Social Network Users?. In *Internet of Things (iThings/CPSCOM), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing*. IEEE, 266–272.
- [25] Delip Rao, David Yarowsky, Abhishek Shreevats, and Manaswi Gupta. 2010. Classifying Latent User Attributes in Twitter. In *Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents*. ACM, 37–44.
- [26] Gerard Salton, Anita Wong, and Chung-Shu Yang. 1975. A Vector Space Model for Automatic Indexing. *Commun. ACM* 18, 11 (1975), 613–620.
- [27] Aaron Smith, Anderson Monica, and Caiazza Tom. 2018. Social Media Update in 2018. (2018).
- [28] Michael A Stefanone, Derek Lackaff, and Devan Rosen. 2010. The Relationship Between Traditional Mass Media and “Social Media”: Reality Television as a Model for Social Network Site Behavior. *Journal of Broadcasting & Electronic Media* 54, 3 (2010), 508–525.
- [29] Barbara G Tabachnick and Linda S Fidell. 2007. *Using Multivariate Statistics*. Allyn & Bacon/Pearson Education.