You Are What You Post: What the Content of Instagram Pictures Tells About Users' Personality

Bruce Ferwerda*
Department of Computer Science and Informatics
School of Engineering
Jönköping University
P.O. Box 1026
SE-551 11, Jönköping, Sweden
bruce.ferwerda@ju.se

Marko Tkalcic
Faculty of Computer Science
Free University of Bozen-Bolzano
Piazza Domenicani 3
I-39100, Bozen-Bolzano, Italy
marko.tkalcic@unibz.it

ABSTRACT
Instagram is a popular social networking application that allows users to express themselves through the uploaded content and the different filters they can apply. In this study we look at the relationship between the content of the uploaded Instagram pictures and the personality traits of users. To collect data, we conducted an online survey where we asked participants to fill in a personality questionnaire, and grant us access to their Instagram account through the Instagram API. We gathered 54,962 pictures of 193 Instagram users. Through the Google Vision API, we analyzed the pictures on their content and clustered the returned labels with the k-means clustering approach. With a total of 17 clusters, we analyzed the relationship with users' personality traits. Our findings suggest a relationship between personality traits and picture content. This allow for new ways to extract personality traits from social media trails, and new ways to facilitate personalized systems.

Author Keywords
Personality, Instagram, picture content, social media

INTRODUCTION
Personality traits have shown to be a useful concept to rely on when considering personalizations of user experiences in a system. This because personality has shown to be a stable construct over time, and reflects the coherent patterning of one's affect, cognition, and desires (goals) as it leads to behavior [22]. The stability and coherency that personality bring, has shown to be useful for systems to infer users' preferences and to provide personalized experiences to users (e.g., [6]). Systems that use personality-based personalizations have shown to have an advantage over systems not using personality information [15]; an advantage is created in terms of increased users' loyalty towards the system and decreased cognitive effort.

The usefulness of personality for personalization is shown in its domain independency: once the personality of users is known, it can be used across domains for personalization [1]. This allows for personality extraction in one domain and implementation in another. Hence, the relationships between personality traits and users' behavior preferences and needs are increasingly being investigated (e.g., health [14, 25], education [3, 19], movies [4], music [6, 8, 5, 7, 11, 26], marketing [20]) in order to learn about the connection between personality traits and specific behaviors.

Since personality traits of users are increasingly being used to provide a personalized experience to users, there is an increased interest in how to implicitly acquire these traits for implementation. A useful source of information are social networking services (SNSs). SNSs are increasingly interconnected with applications through so called "single sign-on buttons" (SSO buttons). [1] The abundance of information that becomes available from the connected SNSs can be used to infer users' personality traits from (e.g., Facebook [7], Twitter [21, 24], and Instagram [9, 10]).

In this work we join the personality extraction research. We specifically focus on Instagram, a popular mobile photo-sharing, and SNS, with currently over 800 million users. [2] With the content as well as with the filters that Instagram allows users to apply to their pictures, users are able to express a personal style and create a seeming distinctiveness. Hence, personality information about users may be hidden in the pictures that users upload to Instagram. Whereas prior work on Instagram focused on the picture properties (i.e., hue, saturation, valence relationship) [9, 10], we focus on the content of the posted pictures on Instagram and explore the relationship with the personality traits of Instagram users. By analyzing the Instagram pictures on their content using the Google Vision

*Also affiliated with the Department of Computational Perception, Johannes Kepler University, Altenberger Strasse 69, 4040, Linz (Austria), bruce.ferwerda@jku.at

HUMANIZE '18, March 11, 2018, Tokyo, Japan

[1] Buttons that allow users to easily register and log in to a system with their social media account.
[2] https://instagram.com/press/ (accessed: 08/12/2017)

API [3], we were able to find distinct correlations between users' personality traits and the content of the pictures they post on Instagram.

RELATED WORK

There is an increasing body of work that looks at how to implicitly acquire personality traits of users. Since all kind of information can relate to personality traits, even information that is not directly relevant for a specific purpose may contain information that is useful for the extraction of personality (e.g., Facebook [7], Twitter [21, 24], and Instagram [9, 10]). The increased connectedness between SNSs and applications through SSO buttons provide an abundance of information that can be exploited to implicitly acquire personality traits of users.

Quercia et al. [21] looked at Twitter profiles and were able to predict users' personality traits by using their number of followers, following, and listed counts. With these three characteristics they were able to predict personality scores with a root-mean-square error 0.88 on a [1,5] scale. Similar work has been done by Golbeck, Robles, and Turner [13] on Facebook profiles. They mainly looked at the sentiment of posted content and were able to create a reliable personality predictor with that information. A more comprehensive work on the prediction of personality and other user characteristics using Facebook likes has been proposed by Kosinski, Stillwell and Graepel [18].

Besides posted content on SNSs, the characteristics of pictures has shown to consist of personality information as well. Celli, Bruni, and Lepri [2] showed that Facebook profile pictures consist of indicators of users' personality. An extension of this work has been recently published [23]. Work of Ferwerda, Schedl, and Tkalcic [12, 10] on Instagram pictures, showed that the way filters are applied to create a certain distinctiveness that can be used to predict personality traits of the poster.

In this work we expand the work of Ferwerda et al. [12, 10] on Instagram pictures. Instead of looking at the picture characteristics (i.e., how filters are applied), we look at the posted content itself.

METHOD

To investigate the relationship between personality traits and picture features, we asked participants to fill in the 44-item BFI personality questionnaire (5-point Likert scale; Disagree strongly - Agree strongly [16]). The questionnaire includes questions that aggregate into the five basic personality traits of the FFM. Additionally, we asked participants to grant us access to their Instagram account through the Instagram API, in order to crawl their pictures.

We recruited 233 participants through Amazon Mechanical Turk, a popular recruitment tool for user-experiments [17]. Participation was restricted to those located in the United States, and also to those with a very good reputation ($\geq$95% HIT approval rate and $\geq$1000 HITs approved) [4] to avoid careless contributions. Several control questions were used to filter out fake and careless entries. This left us with 193 completed and valid responses. Age (18-64, median 30) and gender (104 male, 89 female) information indicated an adequate distribution. Pictures of each participant were crawled after the study. This resulted in a total of 54,962 pictures.

To analyze the content of the pictures, we used the Google Vision API. The Google Vision API uses a deep neural network to analyze the pictures and assign tags ("description") with a confidence level ("score": $r\varepsilon[0,1]$) to classify the content (example given in Listing 1).

```
[{
        "score": 0.8734813,
        "mid": "/m/06__v",
        "description": "snowboard"
}, {
        "score": 0.8640924,
        "mid": "/m/01fklc",
        "description": "pink"
}, {
        "score": 0.81754106,
        "mid": "/m/0bpn3c2",
        "description": "skateboarding
           equipment and supplies"
}, {
        "score": 0.8131781,
        "mid": "/m/06_fw",
        "description": "skateboard"
}, {
        "score": 0.7329241,
        "mid": "/m/05y5lj",
        "description": "sports equipment
           "
}, {
        "score": 0.64866644,
        "mid": "/m/02nnq5",
        "description": "longboard"
}]
```

Listing 1. Example JSON file returned by the Google Vision API for one picture

Using the Google Vision API, we were able to retrieve 4090 unique labels from the Instagram pictures. In order to create an initial clustering of the labels, we used a k-means clustering method that is applied to the vectors that represent the terms in the joint vector space. The vectors were generated with the doc2vec approach using a set of embeddings that are pre-trained on the English Wikipedia [5]. Using this method we collated the labels into 400 clusters. [6] After that, the output of the k-means was manually checked and the clusters were further (manually) collated into similar categories. This resulted into 17 categories representing:

[3] https://cloud.google.com/vision/

[4] HITs (Human Intelligence Tasks) represent the assignments a user has participated in on Amazon Mechanical Turk prior to this study.

[5] https://github.com/jhlau/doc2vec

[6] The k-means clustering method allows for setting a parameter for the number of clusters to be forced. Different number of clusters were tried out. Setting the k-means to automatically define 400 clusters resulted in clusters with least errors in clustering the labels.

1. Architecture
2. Body parts
3. Clothing
4. Music instruments
5. Art
6. Performances
7. Botanical
8. Cartoons
9. Animals
10. Foods
11. Sports
12. Vehicles
13. Electronics
14. Babies
15. Leisure
16. Jewelry
17. Weapons

For each participant, we accumulated the number of category occurrences in their Instagram picture-collection. Since the number of Instagram pictures in each picture-collection is different, we normalized the number of category occurrences to represent a range of $r\varepsilon[0,1]$. This in order to be able to compare users with differences in the total amount of pictures.

RESULTS

We used the Spearman's correlation analysis to analyze the correlations between the picture content categories and personality traits. Alpha levels were adjusted using the Bonferroni correction to limit the chances of a Type I error. The reported significant results adhere to alpha levels of $p < .001$ (see Table 1). Several correlations were found that indicate a higher usage of posting pictures with a certain content depending on personality traits. The correlations between the picture content categories and personality traits are discussed below.

Openness to experience: Openness to experience was found to correlate with the music instruments category (category #4). This shows that those scoring high in the openness to experience trait in general post more pictures consisting of music instruments.

Conscientiousness: A positive correlation was found between conscientiousness and the categories #3 (clothing) and #11 (sports). This indicates that conscientious participants more frequently shared pictures consisting of content with clothing or sports.

	O	C	E	A	N
1	-0.009	-0.009	0.044	-0.002	-0.043
2	-0.039	-0.075	0.023	0.115	0.108
3	0.040	**0.148**	0.110	**0.234**	**-0.184**
4	**0.156**	0.133	0.034	0.049	-0.081
5	0.048	-0.003	0.122	0.111	-0.065
6	0.105	0.113	0.088	0.051	-0.027
7	0.002	-0.034	-0.074	0.099	0.057
8	0.027	-0.040	0.053	0.050	-0.076
9	0.008	-0.003	-0.008	-0.015	0.112
10	-0.069	0.027	-0.012	-0.029	-0.016
11	-0.087	**0.156**	0.023	-0.003	-0.135
12	-0.067	0.054	0.024	0.054	-0.028
13	-0.057	0.097	**0.167**	0.062	-0.132
14	-0.009	0.024	-0.026	0.010	0.058
15	-0.042	0.112	0.085	**0.180**	-0.124
16	-0.055	-0.070	-0.052	-0.017	**0.188**
17	0.009	0.096	-0.019	0.041	0.032

Table 1. Spearman's correlation between picture content categories and personality traits. Significant correlations after Bonferroni correction are shown in boldface ($p <$.001).

Extraversion: We found a correlation between electronics (category #13) and extraversion. Extraverts tend to post pictures on their Instagram account consisting of electronics.

Agreeableness: Positive correlations were found between agreeableness and the the categories #3 (clothing) and #15 (leisure). This means that the Instagram picture-collections of agreeable participants consist of pictures with clothing or leisure content.

Neuroticism: A negative correlation was found with category #3 (clothing) and a positive correlation was found with category #16 (jewelry) and those scoring high on neuroticism. The results show that people who score high on neuroticism tend to have less pictures with clothing content, but in general have more content with jewelry.

CONCLUSION AND OUTLOOK

We found the content of Instagram picture features to be correlated with personality. A summary of the correlations between the picture content and personality traits can be found in Table 2.

Personality	Picture content
Openness to experience	Music instruments
Conscientiousness	Clothing, sports
Extraversion	Electronics
Agreeableness	Clothing, leisure
Neuroticism	Clothing (-), jewelry

Table 2. Interpretation and summary of the correlations found between personality traits and picture properties. Unless indicated with "(-)," the results indicate positive correlations. The content correlations apply for the pictures of participants who score high in the respective personality trait.

The identification of the correlations between image categories and user personality is the first step towards unobtrusive personality detection and personalization. In future work we plan to use the automatically detected categories as features for the unobtrusive prediction of personality using machine learning techniques. With this work we are complementing prior work of Ferwerda et al. [12, 10] in which they used the picture properties of Instagram pictures to find relations with personality traits as well creating a predictive model of personality traits. Future work will focus on combining the relevant picture features of prior work with the categories that we laid out in this work to improve the predictive models that can be created for personality prediction.

ACKNOWLEDGEMENTS

We would like to thank Marcin Skowron for his help and expertise on processing the data into clusters.

REFERENCES

1. Iván Cantador, Ignacio Fernández-Tobías, and Alejandro Bellogín. 2013. Relating personality types with user preferences in multiple entertainment domains. In *CEUR Workshop Proceedings*. Shlomo Berkovsky.
2. Fabio Celli, Elia Bruni, and Bruno Lepri. 2014. Automatic personality and interaction style recognition from facebook profile pictures. In *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 1101–1104.
3. Guanliang Chen, Dan Davis, Claudia Hauff, and Geert-Jan Houben. 2016. On the impact of personality in massive open online learning. In *Proceedings of the 2016 conference on user modeling adaptation and personalization*. ACM, 121–130.
4. Li Chen, Wen Wu, and Liang He. 2013. How personality influences users' needs for recommendation diversity?. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 829–834.
5. Bruce Ferwerda, Mark Graus, Andreu Vall, Marko Tkalcic, and Markus Schedl. 2016. The influence of users' personality traits on satisfaction and attractiveness of diversified recommendation lists. In *4 th Workshop on Emotions and Personality in Personalized Systems (EMPIRE) 2016*. 43.
6. Bruce Ferwerda and Markus Schedl. 2014. Enhancing Music Recommender Systems with Personality Information and Emotional States: A Proposal.. In *UMAP Workshops*.
7. Bruce Ferwerda and Markus Schedl. 2016. Personality-Based User Modeling for Music Recommender Systems. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 254–257.
8. Bruce Ferwerda, Markus Schedl, and Marko Tkalcic. 2015a. Personality & Emotional States: Understanding Users' Music Listening Needs.. In *UMAP Workshops*.
9. Bruce Ferwerda, Markus Schedl, and Marko Tkalcic. 2015b. Predicting personality traits with instagram pictures. In *Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems 2015*. ACM, 7–10.
10. Bruce Ferwerda, Markus Schedl, and Marko Tkalcic. 2016. Using instagram picture features to predict users' personality. In *International Conference on Multimedia Modeling*. Springer, 850–861.
11. Bruce Ferwerda, Marko Tkalcic, and Markus Schedl. 2017. Personality Traits and Music Genres: What Do People Prefer to Listen To?. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. ACM, 285–288.
12. Bruce Ferwerda, Emily Yang, Markus Schedl, and Marko Tkalcic. 2015. Personality traits predict music taxonomy preferences. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2241–2246.
13. Jennifer Golbeck, Cristina Robles, and Karen Turner. 2011. Predicting personality with social media. In *CHI'11 extended abstracts on human factors in computing systems*. ACM, 253–262.
14. Sajanee Halko and Julie A Kientz. 2010. Personality and persuasive technology: an exploratory study on health-promoting mobile applications. In *International Conference on Persuasive Technology*. Springer, 150–161.
15. Rong Hu and Pearl Pu. 2009. Acceptance issues of personality-based recommender systems. In *Proceedings of the third ACM conference on Recommender systems*. ACM, 221–224.
16. Oliver P John and Sanjay Srivastava. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research* 2, 1999 (1999), 102–138.
17. Aniket Kittur, Ed H Chi, and Bongwon Suh. 2008. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 453–456.
18. Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America* 110, 15 (mar 2013), 5802–5. DOI: http://dx.doi.org/10.1073/pnas.1218772110
19. Michael J Lee and Bruce Ferwerda. 2017. Personalizing online educational tools. In *Proceedings of the 2017 ACM Workshop on Theory-Informed User Modeling for Tailoring and Personalizing Interfaces*. ACM, 27–30.
20. S C Matz, M Kosinski, G Nave, and D J Stillwell. 2017. Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences* 114, 48 (nov 2017), 12714–12719. DOI: http://dx.doi.org/10.1073/pnas.1710966114

21. Daniele Quercia, Michal Kosinski, David Stillwell, and Jon Crowcroft. 2011. Our Twitter profiles, our selves: Predicting personality with Twitter. In *Proceedings of the International Conference on Social Computing (SocialCom)*. IEEE, 180–185.

22. William Revelle. 2009. Personality structure and measurement: The contributions of Raymond Cattell. *British Journal of Psychology* 100, S1 (2009), 253–257.

23. Cristina Segalin, Fabio Celli, Luca Polonio, Michal Kosinski, David Stillwell, Nicu Sebe, Marco Cristani, and Bruno Lepri. 2017. What your Facebook Profile Picture Reveals about your Personality. In *Proceedings of the 2017 ACM on Multimedia Conference, MM 2017, Mountain View, CA, USA, October 23-27, 2017*. 460–468. DOI:http://dx.doi.org/10.1145/3123266.3123331

24. Marcin Skowron, Marko Tkalčič, Bruce Ferwerda, and Markus Schedl. 2016. Fusing social media cues: personality prediction from twitter and instagram. In *Proceedings of the 25th international conference companion on world wide web*. International World Wide Web Conferences Steering Committee, 107–108.

25. Kirsten A Smith, Matt Dennis, and Judith Masthoff. 2016. Personalizing reminders to personality for melanoma self-checking. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*. ACM, 85–93.

26. Marko Tkalčič, Bruce Ferwerda, David Hauger, and Markus Schedl. 2015. Personality correlates for digital concert program notes. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 364–369.