Inferring the Relationship between Users' Personality and their image on Flicker using deep learning and clustering

Rojiar Pir Mohammadiani
Computer engineering department, university of Kurdistan
Sanandaj, Iran
R.pirmohamadiani@uok.ac.ir

Hamid Sadeghi
Computer engineering department, university of Wayne
Michigan, USA
hamid.sadeghi@wayne.edu

Abstract—Most of the social media such as Instagram/Facebook/Flicker/Twitter allow users to express themselves through the post content, like content and profile information. Each of these activities is a reflection of a user's personality. Images are currently present in a significant fraction of data in social media. Therefore, these days personality trait prediction by image-based content analysis attract more attention than text-based analysis among researchers. However, investigating of large amount of unstructured data of presented images is complicated. Therefore, proposing models that can prepare this knowledge briefly and structured is crucial. In this paper has been attempted to detect personality by combined deep learning and clustering approach that accurately inferring relationship between users' images content and their personality in Flicker. We considered different modalities of interaction including: posts, likes and their profile images in this platform. The results of this paper will be useful in different areas such as marketing advertisement, psychology and sociology.

Keywords—Personality trait; social media; Image content; Deep learning; clustering

I. INTRODUCTION

The enlargement employment of social media motivates users to utilize social medias and do some related activities such as posting content, liking others' posts, or making a user profile. Since, user generated content in social media platforms is important to understand how users reveal themselves across multiple activities. There are many researches to develop new ways of measuring and modeling human behavior in online social platforms such as evaluating influence level of users, evaluating knowledge level of users in terms of expertise and novelty [13, 14]. But, one of the most interesting area is studying personality traits using big five model [4]

Personality a user defines as collection of feeling pattern, preferences, and thinking [1]. Based on the psychology knowledge personality differences can be organized along five broad dimensions – Extraversion, Agreeableness, Neuroticism, Conscientiousness, and Openness-to-Experience (with the useful acronym, OCEAN). This model known as "Big five" [4]. The user behavior in social media has been explored to understand more about the relationship between personality

traits and behavior of users. To this regard, some researches have been conducted to gain an understanding of this relationship in different areas (e.g., health [8], education, movies, and music [2]). Consequently, personality traits information has been utilized for boosting applications such as using in marketing [7].

Many studies have been conducted on relationship between user's personality traits using big five model and their modalities of interaction on dissimilar social media platforms.

Some studies have been considered a single platform such as liked pages of Facebook [20], content of liked and posted images on twitter [17] gallery and liked on Flickr [18, 21], however some of studies have been examined cross-platform [11]

In preliminary studies, it is an ordinary way to find out users' personality traits by their texts in social media. Feature extraction have been utilized to judge user personality based on the employment of words in dissimilar semantic categories [9, 10, 12]. Fewer studies have contemplated other features, such as user likes and profile images [5, 6, 11]. These days, most of the social media such as Instagram/Facebook/Flicker/Twitter acknowledge users to reveal themselves through the post or like images. Images are currently present in a notable portion of data in social media [20]. Therefore, personality information about users may be hidden in the pictures. So, these days personally trait prediction by image-based content analysis catch more attention than text-based analysis among researchers. However, little is known about how images content is related to the with their users' personality. For instance,

Which image features characterize each personality type?

How personality traits can impact what users post, like and select as profile picture?

Current deep learning algorithms enable us to get different concepts from images which is useful to analyze the relationship of posted images and personality trait of the people how post them. In this study, we extracted the concept of images in the form of a collection of objects and places using convolutional neural network. We have considered all the

photos related to each user as a three different modalities type including: posts, likes and their profile images in order to have a more comprehensive view of the users. However, investigating of large amount of unstructured data of presented images is complicated. Therefore, proposing models that can prepare this knowledge briefly and structured is crucial.

One of the useful algorithms in the field of machine learning for summarizing data is clustering. Therefore, in this paper, the clustering algorithm is used in two parts. In the first part, users are divided into 4 groups according to the similarity in the content of their photos (termed as 'content clustering' in the rest of this paper). In the second part, due to the fact that the number of features extracted from the images is very high, these features are categorized to 15 groups according to semantic similarity (termed as 'semantic clustering' in the rest of this paper). Then, we present a comprehensive analysis that focuses on the types of relationships between the personally trait and content of images using users clustering. Formalizing and finding out the relationship between the two modalities – personally trait and images– is useful in several areas such as psychologists and marketing.

II. LITERATURE REVIEW

Big Five is one of the examples of a trait theory. The 5 main personality is compromise of openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. In table 1 personality features along with personality details has been demonstrated.

Table1. Personality features of each personality trait

Personality trait	Sample and positive side descriptions of the corresponding feature	Sample and negative side descriptions of the corresponding feature
Extraversion	Outgoing, courageous, active, energetic, eager	Quiet, lethargic,
Agreeableness	Cooperative, considerate, empathic, generous , polite and kind	Aggressive, nude, spiteful, stubborn
Neuroticism	Emotionally stable, ambisious	Worries a lot, easily frustrated and insecure in relationship
Consciousness	Does a thorough job and future-oriented	unreliable, Tends to be lazy, careless,
Openness	Imaginative, Inventive	Simple without imagination

As a matter of knowing user's personality type, we need to find an automatic personality detection model according to features we extract from user's profile. These features can be classified into lingual features, extracted from the textual content available in user's profile such as status, interests, about me text, and so on, and non- lingual features define as set of behaviors, attributes and so on [3]. In regards to textual features, most studies include LIWC or MRC features in which they examined lingual styles to find out language use as an individual distinction. Pennebaker and King [12] as well as Mairesse et al. [9] examined and realized correction between big five personality traits and MRC features.

Yezheng, et al by using Facebook data sets could detect personality trait as a problem categorizing multiple layers. With this assumption that subjects by binomial distribution on words and gaussian distribution of personality traits models. Some assessment shows that model is interpretable. In addition, it states that personality trait is a powerful tool for expressing explanation and foreseeing user's behavior [25].

Farandi and his collogues worked on Twitter and Facebook data. In their work, user's personality made from content which made by users in social media and by using LIWC and three machines leaning algorithms SVM, decision tree and regression. They found out feature selection not necessarily by using from related increase performance of the user. However, by reducing size of desired space can increase performance of algorithms.

Pierre et al. present a new model for predicting personality in social media. They utilized Twitter as a sample data and by combining Word Embedding features with gaussian processes regression present their study. Result from this approach indicates that accuracy increased in compare to earlier studies and additionally, mean coloration 33 presents was lager on big five categories with proposed features. This approach reveals that 3gram features with RR get better result in compare with LIWC features [26]

Kosinski et al. worked Facebook likes. They used their model to demonstrate how they can predict sensitive personality attributes by user's like on Facebook [20]

Recently because of growing use of social Media, utilization of images on social media growingly has been increased [3]. Most of lately studies on images concentrate simply on profile images by using facial features. For instance, a research has been conducted on Facebook to predict user's personalities based on their profile images with 65 percent accuracy [6]. In another study, a combination of a face recognition data sets and online randomly selected portraits led to model user's personality [3]. Lately, a study has been conducted in which aesthetic features and facial features were utilized to predict personality on twitter.

Guntuku et al. showed how content of posted and liked images on Twitter can help to predict personality trait of users [17].

Segalin et al. used a dataset contains Flickr liked images and self-assessed for each user. They trained their model using the dataset and could predict big five personality trait based on that [21]

Riahi Samani et al. worked on two platforms (Flickr and Twitter). They did their prediction based on a model that was trained on thier dataset. They consider different modalities of interaction [11].

None of the latest studies considered the relationship between users which participate in their studies and features. In our study we first find four groups of user clusters, then we could find 15 clusters in our features list and finally, find out relationship between these fifteen clusters and four group of users.

III. PROPOSED METHOD

A. Proposed methods

Figure 1 presents the architecture of our system, which consists of five major components: 1) data collection, 2) feature engineering, 3) clustering users based on their similarity in their image content (content clustering) 4) features extracted from the images clustering according to semantic similarity by using GloVe algorithm (semantic clustering). c algorithm is an unsupervised algorithm [23] 5) Inferring the relationship between image content and user personality.
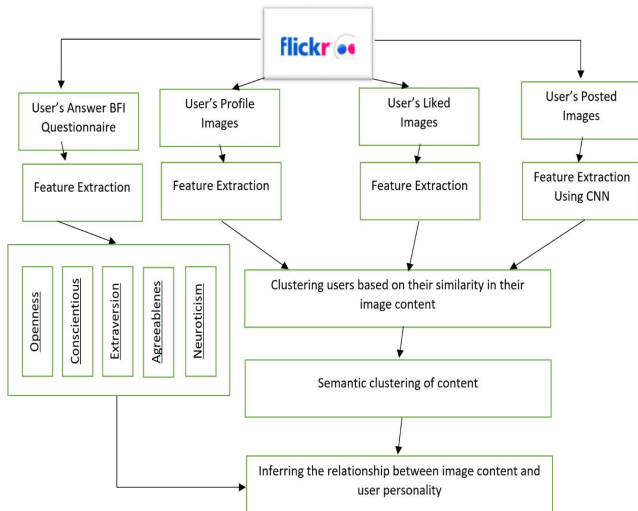
Fig. 1. Proposed methods

B. GloVe algorithem

In this part, we described how we used GloVe algorithm to cluster our features. Generally, we have 1365 features (1000 objects and 365 scene categories). Our result shows feature column number (1-1365), column value (feature name) and semantic clustering number. Fig.2 shows how 1365 features dived into 15 semantic groups
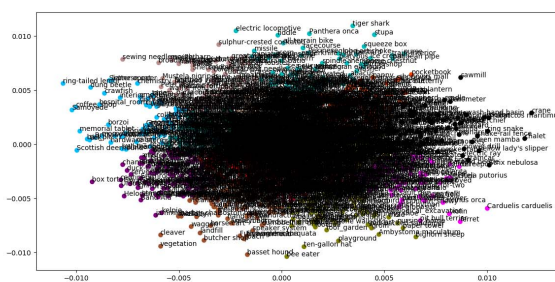

Fig.2. Result of semantic clusters using GloVe

C. Data

In order to evaluate the relationship between picture features and personality traits we asked our participants to fill out a BFI personality questionnaire. Our questionnaire is a five-point scale questionnaire. The questionnaire involves questions that aggregate into the five basic personality traits as

we explained earlier. Additionally, we use the Flickr API (https://www.flickr.com/services/api/) to extract profile images of those users and their posted and liked images. Table 2 shows the descriptive statistics of data set.

Table 2. Descriptive statistics of data set

Modality	Total number	Total number per user
Posted images	72997	274
Liked (fave) images	60001	203
Profile images	295	1

In data collection component, a data set contains a set of self-assessed and perceived personalities for 295 random users from Flickr was collected. We use this data set for inferring the relationship between user personality and different interactions

D. Feature engineering

In feature engineering layer after the preprocessing, to analyze content of the images on Flicker, CNN algorithm through ImageNet was utilized.

In our work we used a CNN model which is called VGGnet and have been proposed by K.simonyan et al. This model is a 16-19 hidden layers deep convolutional neural network which challenge [15]. In addition, we utilized ImageNet data set which consist of over 15 million images. The VGGNet takes 224x224 RGB image as input image and then the image passed through a stack convolution layers.

For profile images, we use the features extracted from the profile image of the user and for liked and posted images we perform a mean feature pooling of all liked and posted images each across all images per user. Moreover, we have utilized the model trained on 1000 object ImageNet tagset as well as 365 standard sense sets. In addition to the model, the prediction probabilities across all object group and scene have been applied as features [11].

E. Users clustering based on their similarity in their image content

However, investigating of large amount of unstructured data of presented images is complicated. Therefore, proposing models that can prepare this knowledge briefly and structured is crucial. Then, users based on their correspondence in their image was clustered using k-means algorithm. For each user, we add the cluster number to which it belongs as a new feature in the table that shows the user personality traits of that user. Then, the average personality of users who belong to each cluster was calculated. According to chart1 we can understand following consequences:

- Dominant personality type in group one is Openness

- Dominant personality type in group two is Conscientiousness

- Dominant personality type in group three is Agreeableness

- Group four tend to Openness and Agreeableness personality type

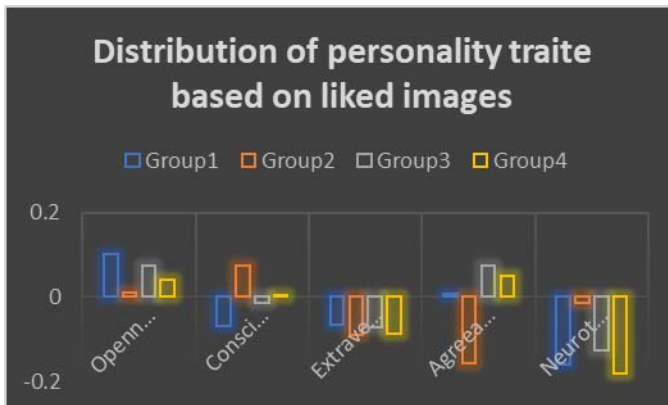- Neuroticism is not significant or not much considerable in all groups

Chart 1. Distribution of personality traite based on liked images

According to chart2 we can understand following results:

- Dominant personality type in group one is Openness

- Dominant personality type in group two is Openness

- Dominant personality type in group four is Openness and then Agreeableness

- Dominant personality type in group three is Openness and then Agreeableness

- Dominant personality type in group four is Openness and Agreeableness

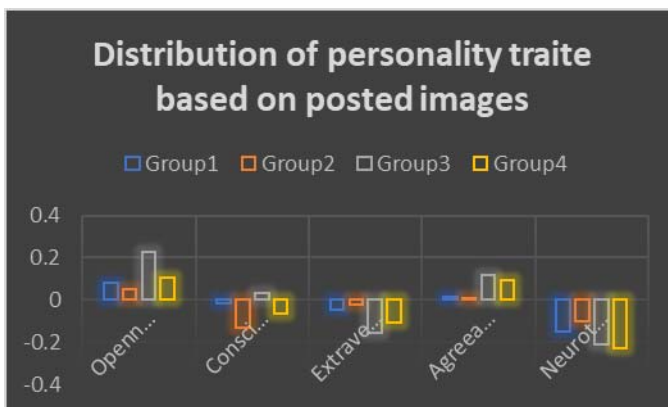- Group three more tend to Openness personality type in compare to other groups

Chart 2. Distribution of personality traite based on posted images

According to chart3 we can understand following consequences:

- Dominant personality type in group one is Openness and then Agreeableness

- Dominant personality type in group two is Openness and then Neuroticism

- Dominant personality type in group three is Agreeableness and then Extraversion

- Group four more tend to Extraversion personality type in compare to other groups

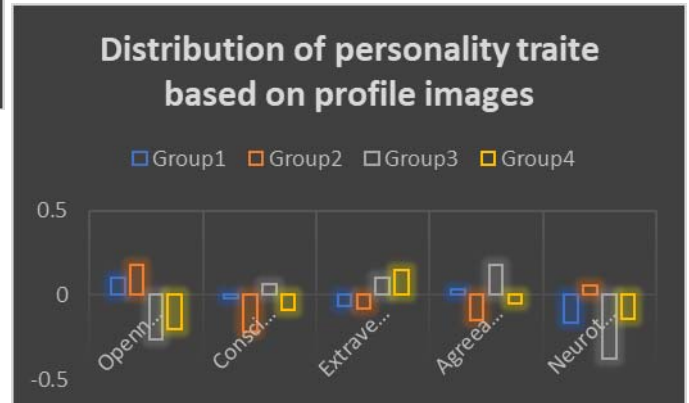- Group three more tend to Agreeableness personality type in compare to other groups

Chart 3. Distribution of personality traite based on profile images

F. Semantic clustering content

A k-means clustering has been utilized to make an initial clustering which has been applied to vectors that express the terms in the space of joint vector. GloVe approach and a set of pre-trained data on the English Wikipedia has been used to generate the vectors. K-means clustering method enables us to put a parameter for cluster number. Different number of clusters has been utilized to ensure we are using the best number. Assigning the k-mean to 40 clusters infers clusters with least error in clustering of the labels. As the next step, we manually checked the output of the k-means and then divided into similar categories. This led to 15 distinct categories

For each participant, we accumulated the number of category occurrences in their Flicker picture-collection. This in order to be able to compare users with differences in the total amount of pictures. Finally, we analyze the correction between personality traits and the content of images.

G. Results

Here we used four groups of people that we got in content clustering we called them group one to four in content clustering part.

According to chart 4 we can conclude following results:

- Users belong to group4 of content clustering groups mainly published contents belongs to

group 15 of semantic clustering. Users in group4 have Openness and Agreeableness personality types from chart 2 which means Openness personality type published contents of group 15 more than any of 15 semantic groups.

- Users in group3 of content clustering mainly published contents belongs to group 4 of semantic clustering. as we understood in chart 2, personality type group 3 of content clustering is Openness and then Agreeableness which means Openness and Agreeableness personality types published content of group 4 of semantic groups

- Users in group2 of content clustering mainly published contents belongs to group15 ,12, 8 and 6 more than other groups. Based on chart 2, we can infer that Openness published contents of groups 15, 12, 8 and 6

- Finally, we can infer users in group1 of content clustering almost published content of all 15 groups of semantic clustering. According to chart 2, personality type group 1 is Openness. And published contents of all 15 groups of semantic clustering.
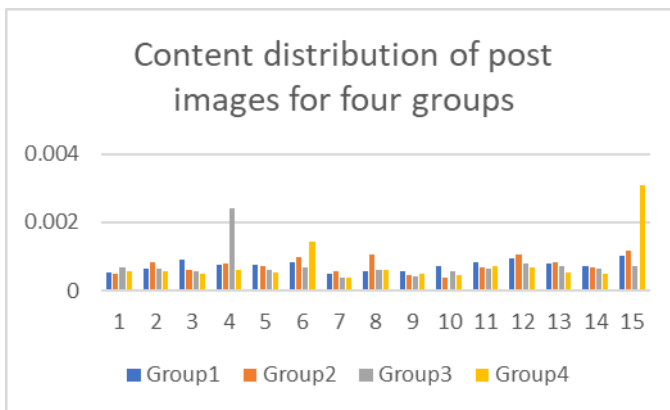

Chart 4. Content distribution of post images for four groups

According to chart 5 we can infer following results:

- Users belong to group 2 of content clustering mainly published contents belongs to group 12 and 10 of semantic clustering. Users in group 2 have Consciousness and Openness personality type from chart 1 which means Consciousness and Openness personality type published group 12 and 10 content more than any of 15 semantic groups.

- Users in group 4 of content clustering mainly published contents belongs to group 15 and then group 6. as we understood from chart 1, personality type group 4 of content clustering is Agreeableness and then Openness which means Agreeableness and Openness personality types

published contents of group 15 and then 6 more than any of 15 semantic groups.

- Users belong to group 1 of content clustering mainly published contents belongs to group 15 of semantic clustering. Users in group 1 have Openness personality type from chart 1 which means Openness personality type published group 15 content more than any of 15 semantic groups.

- Finally, we can infer users in group 3 of content clustering almost published contents of all 15 groups of semantic clustering. According to chart 1, personality type group 3 is Openness and Agreeableness.
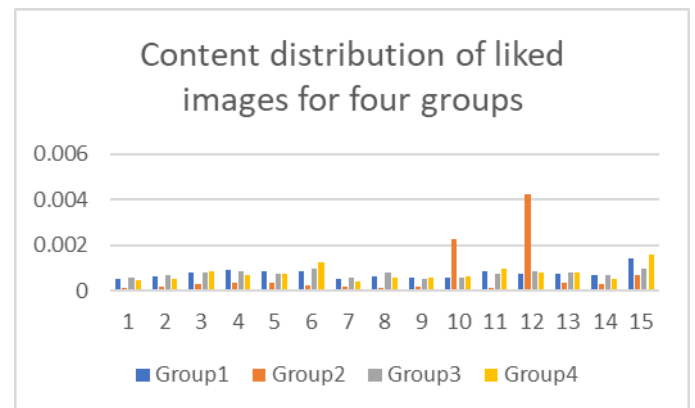

Chart 5. Content distribution of liked images for four groups

According to chart 6 we can infer following results:

- Users in group 2 of content clustering mainly published contents belongs to group 12 and then group 10. As we understood from chart 3, personality type group 2 of content clustering is Openness and then Neocriticism which means Openness and Neocriticism personality types published contents of group 12 and then group 10 semantic groups.

- Users in group 4 of content clustering mainly published contents belongs to group 15 and then group 6. As we understood from chart 3, personality type group 4 of content clustering is Extraversion which means Extraversion personality type published content of group 15 and then group 6 semantic groups.

- Users in group 1 of content clustering mainly published contents belongs to group 15. As we understood from chart 3, personality type group 1 of content clustering is Openness and Agreeableness which means Openness and Agreeableness personality type published contents of group 15 semantic groups.

- Finally, we can infer users in group 3 of content clustering almost published content of all 15 groups of semantic clustering. According to chart 3, personality type group 3 is Agreeableness.
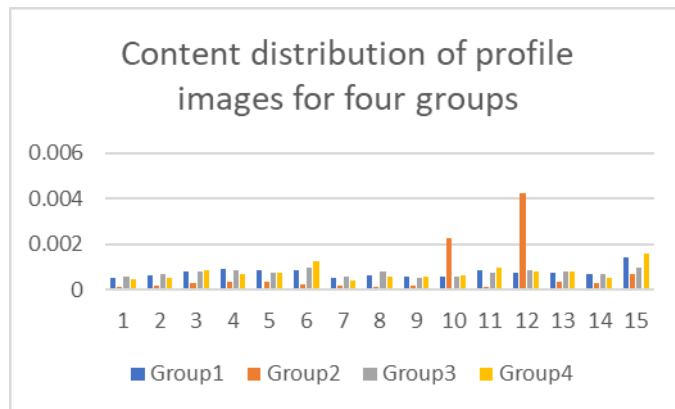

Chart 6. Content distribution of profile images for four groups

H. Conclusions

The data available on social media such as Facebook, Twitter and Instagram can be useful to detect user's personality traits. The model we used in this work enables us to predict how users express themselves across different images content on Flicker. Our findings offer a relationship between personality traits and picture content. This allows for new ways to extract personality traits from social media trails, and new ways to facilitate personalized systems or like predictions.

As future work we recommend combination of the relevant images features in other studies with what we have reached in this study to enhance the predictive models which will be made in future study to predict personality traits.

References

[1] L. Robins, A. Pervin (Eds.). Handbook of personality: Theory and research (3rd ed.), pp. 421- 447). New York, NY, US: Guilford Press.

[2] I. Cantador, I. Fernández-Tobías, A. Bellogín, " Relating personality types with user preferences in multiple entertainment domains," In: CEUR Workshop Proceedings, vol. 997, 2013.

[3] V. Kaushal, and M. Patwardhan. "Emerging Trends in Personality Identification Using OnlineSocial Networks—A Literature Survey", ACM Transactions on Knowledge Discovery from Data, 2018.

[4] R.R. McCrae, and J. OP, "An Introduction to the Five-factor Model and its Applications," Journal of Personality, vol. 60, 1992. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

[5] H. Wei, F. Zhang, N. Jing Yuan, Ch. Cao, H. Fu, X. Xie, Y. Rui, and Wei-Ying Ma. "Beyond the Words", Proceedings of the Tenth ACM International Conference on Web Search and Data Mining - WSDM '17, 2017

[6] H. Jun, L. Shuai, Sh. Jinming, L. Yue, W. Jingwei, and J. Peng. "Facial Expression Recognition Based on VGGNet Convolutional Neural Network", Chinese Automation Congress (CAC), 2018.

[7] M. Tkalčič, B. De Carolis, M. de Gemmis, and A. Košir, "Emotions and Personality in Personalized Systems ", Boston, MA, USA,EMPIRE 2016.

[8] T. W. Smith, J. MacKenzie, A. Rev, "Personality and risk of physical illness," Clin. Psychol. Vol. 2, pp. 435 467, 2006.

[9] Mairesse and M. Walker, "PERSONAGE: Personality generation for dialogue," In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL'07), pp. 496–503, 2007.

[10] Y. Hernández, C. Acevedo Peña, A. Martínez. "Chapter 17 Model for Personality Detection Based on Text Analysis", Springer Science and Business Media LLC, 2018.

[11] Z. Riahi Samani, Sh. Chandra Guntuku, M. Ebrahimi Moghaddam, and D. Preoţiuc-Pietro, "Cross-platform and cross-interaction study of user personality based on images on Twitter and Flickr", PLOSONE, 2018.

[12] J. W. Pennebaker and L. A. King, "Linguistic styles: Language use as an individual diference," Journal of Personality and Social Psychology, vol. 77, pp. 1296–1312, 1999.

[13] W. Chung, and D. Zeng, "Dissecting emotion and user influence in social media communities: An interaction modeling approach", Information & Management,Vol. 57 (1), 103108, 2020.

[14] X. Zhou, B. Wu, and Q. Jin, " User role identification based on social behavior and networking analysis for information dissemination", Future Generation Computer Systems,Vol. 96, pp. 639-648, 2019.

[15] K. Simonyan, and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409, 1556, 2014.

[16] G. Appel, L. Grewal, and R. Hadi, " The future of social media in marketing". J. of the Acad. Mark. Sci," vol. 48, pp. 79–95, 2020. https://doi.org/10.1007/s11747-019-00695-1

[17] Guntuku, S. C.; Lin, W.; Carpenter, J.; Ng, W. K.; Ungar, L. H.; and Preot¸iuc-Pietro, D. 2017a. Studying personality through the content of posted and liked images on twitter. In Proceedings of the 2017 ACM on web science conference, 223–227. ACM

[18] Segalin, C., Perina, A., Cristani, M., Vinciarelli, A.: The pictures we like are our image: Continuous mapping of favorite pictures into self-assessed and attributed personality traits. IEEE Trans. Affective Comput. pp. 1–1 (2016), http://dx.doi. org/10.1109/TAFFC.2016.2516994

[19] Guntuku, S. C.; Zhou, J. T.; Roy, S.; Weisi, L.; and Tsang, I. W. 2016. Who likes what, and why? insights into personality modeling based on imagelikes'. IEEE Transactions on Affective Computing.

[20] Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences, USA, 110, 5802–5805. doi:10.1073/pnas.1218772110

[21] C. Segalin, D. S. Cheng, M. Cristani, Social profiling through image understanding: Personality inference using convolutional neural networks, Computer Vision and Image Understanding.

[22] H. Zhu, L. Li, H. Jiang, and A. Tan, "Inferring personality traits from attentive regions of user liked images via weakly supervised dual convolutional network," in Neural Processing Letters, 2019, doi:10.1007/s11063- 019-09987-7.

[23] http://text2vec.org/GloVe.html

[24] Y. Zhang and B. Wallace, "A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification," arXiv preprint arXiv:1510.03820, 2015.

[25] Yezheng Liu, Jiajia Wang &Yuanchun Jiang.(2015)." PT-LDA: A Latent Variable Model to Predict Personality Traits of Social Network Users.":Neurocomputing, 19 October: www.elsevier.com/locate/neucom

[26] Pierre-Hadrien Arnoux, et al.(2017)." 25 Tweets to Know You: A New Model to Predict Personality with Social Media":Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved