2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

The Role of User Profiles for Fake News Detection

Kai Shu*, Xinyi Zhou† Suhang Wang‡, Reza Zafarani†, and Huan Liu*
*Arizona State University, {kai.shu, huan.liu}@asu.edu
†Penn State University, szw494@psu.edu
‡Syracuse University, {zhouxinyi, reza}@data.syr.edu

***Abstract*—Consuming news from social media is becoming increasingly popular. Social media appeals to users due to its fast dissemination of information, low cost, and easy access. However, social media also enables the widespread of *fake news*. Due to the detrimental societal effects of fake news, detecting fake news has attracted increasing attention. However, the detection performance only using news contents is generally not satisfactory as fake news is written to mimic true news. Thus, there is a need for an in-depth understanding on the relationship between *user profiles* on social media and fake news. In this paper, we study the problem of *understanding* and *exploiting* user profiles on social media for fake news detection. In an attempt to understand connections between user profiles and fake news, first, we measure users' sharing behaviors and group representative users who are more likely to share fake and real news; then, we perform a comparative analysis of explicit and implicit profile features between these user groups, which reveals their potential to help differentiate fake news from real news. To exploit user profile features, we demonstrate the usefulness of these user profile features in a fake news classification task. We further validate the effectiveness of these features through feature importance analysis. The findings of this work lay the foundation for deeper exploration of user profile features of social media and enhance the capabilities for fake news detection.**

I. Introduction

Due to the increasing amount of time spent on social media, people increasingly tend to seek out and receive their news through social media sites. In December 2016, the Pew Research Center announced that approximately 62% of US adults get news from social media in 2016, while in 2012, only 49% reported reading news on social media.[1] This rapid rate of increase in user engagements with online news can mainly be attributed to the cheap, mobility, and fast dissemination of social media platforms. However, despite these advantages, the quality of news on social media is considered lower than that of traditional news outlets. Every, large volumes of fake news, i.e., news stories with intentionally false information [1], [2], are widely spread online. For example, a report estimated that over 1 million tweets were related to the fake news story "Pizzagate"[2] by the end of 2016 presidential election. Thus it is critical to detect fake news on social media for social good.

However, detecting fake news on social media presents unique challenges. First, fake news is intentionally written to mislead readers, which makes it nontrivial to detect simply based on content; Second, social media data is large-scale, multi-modal, mostly user-generated, sometimes anonymous and noisy. Recent research advancements aggregate uses profiles and engagements on news pieces to help infer articles that are incredible [3], leading to some promising early results. However, no principled study is conducted on characterizing the profiles of users who spread fake/real news on social media. In addition, there has been no research that provides a systematic understanding of (i) what are possible user profile features; (ii) whether these features are useful for fake news detection; and (iii) how discriminative these features are. To give a comprehensive understanding, we investigate the following three research questions:

- **RQ1:** *Which users are more likely to share fake news or real news?*
- **RQ2:** *What are the characteristics/features of users that are more likely to share fake/real news, and do they have clear differences?*
- **RQ3:** *Can we use user profile features to detect fake news and how?*

By investigating **RQ1**, we identify users who are more likely to share fake or real news, which can be treated as representative user sets to characterize user profiles. By answering **RQ2**, we further provide guidance on assessing whether the profiles of identified users are different or not, and to what extent and in what aspects they are different. In addition, by studying **RQ3**, we explore different ways to model user profile features, analyze the importance of each feature and show the feature robustness to various learning algorithms. By answering these research questions, we made the following contributions:

- We study a novel problem of understanding the relationships between user profiles and fake news, which lays the foundation of exploiting them for fake news detection;
- We propose a principled way to characterize and understand user profile features. We perform a statistical comparative analysis of these profile features, including explicit and implicit features, between users who are more likely to share fake news and real news, and show their potentials to differentiate fake news; and
- We demonstrate the usefulness of the user profile features to classify fake news, whose performance consistently outperforms existing state-of-the-art features extracted from news content. We also show that the extracted user profile features are robust to different learning algorithms, with an average $F1$ above 0.90. We further validate the effectiveness of these features through feature importance analysis, and found that implicit features, e.g., political bias, perform better than explicit features.

II. Assessing Users' Sharing Behaviors

We investigate **RQ1** by measuring the sharing behaviors of users on social media on fake and real news. We aim

[1]http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/

[2]https://en.wikipedia.org/wiki/Pizzagate_conspiracy_theory

ASONAM '19, August 27-30, 2019, Vancouver, Canada

ACM ISBN 978-1-4503-6868-1/19/08?/$15.00
http://dx.doi.org/10.1145/3341161.3342927

TABLE I: The statistics of FakeNewsNet dataset

Platform	Politifact	Gossipcop
# Users (without filtering bots)	159,699	209,930
# Sharing	271,462	812,194
# True news	361	4,513
# Fake news	361	4,513

to identify users who are more likely to share fake or real news, which can be further used to characterize discriminative features for fake news detection.

A. Datasets

We utilize one fake news benchmark data repository FakeNewsNet [4]. The datasets are collected from two fact-checking platforms: *Politifact*[3] (**P** for short) and *Gossipcop*[4] (**G** for short), both containing news content with labels annotated by professional fact-checkers, and social context information. News contents include meta attributes (e.g., body text), and social context includes the related user social engagements of news items (e.g., user posting/sharing news) on Twitter. The detailed statistics of the datasets are shown in Table I.

B. Filtering Bot Users

Social bots have played an important role to spread fake news on social media [5]. To alleviate the effects of social bots, we apply one of the state-of-the-art bot detection tool *Botometer*[5] [5] to filter out bot accounts. Botometer takes a Twitter username as an input and utilizes various features extracted from meta-data obtained from Twitter API and outputs a probability in $[0, 1]$, indicating how likely the user is a social bot. Following the common setting, we filter out those users who have a score greater than 0.5. We keep the remaining users and treat them as authentic human users.

C. Identifying User Groups

We identify different subsets of users based on their sharing behaviors on fake and real news. By finding these groups, we want to build representative user sets that are more likely to share fake/real news from which we can further compare the degree of the differences of their profiles to find useful profile features. Towards answering **RQ1**, we adopt the proposed measures in [6], absolute measure and relative measure, to select the top-K users.

Based on the two measures, we introduce a principled way to identify representative user groups $\mathcal{U}^{(f)}$ and $\mathcal{U}^{(r)}$. First, we divide all users into three subsets: (1) "Only Fake": users who only spread fake news; (ii) "Only Real": users who only spread real news; and (iii) "Fake and Real": users who spread both fake and real news Second, we empirically select top 10,000 users from "Only Fake" and "Only Real" ranked by the number of fake news or real news they share; and then we further select users with lower FR scores ($FR \in [0, t]$) and add to $\mathcal{U}^{(r)}$ with a threshold t; and select users with higher FR scores ($FR \in [1-t, 1]$) and add them to $\mathcal{U}^{(f)}$. By changing the threshold of t, we can obtain consistent results when $t < 0.4$, and when $t \geq 0.4$ more noisy tend to be included, and the comparison analysis may not be accurate. Thus we set $t = 0.2$ to reduce the noise for the feature analysis. The selected users are equally sampled for both $\mathcal{U}^{(f)}$ and $\mathcal{U}^{(r)}$.

[3]https://www.politifact.com/

[4]https://www.gossipcop.com/

[5]https://botometer.iuni.iu.edu

III. Understanding User Profiles

We collect and analyze user profile features from different aspects, i.e., *implicit* and *explicit*. Implicit features are not directly available but are inferred from user meta information or online behaviors, such as historical tweets. Explicit features are obtained directly from meta-data returned by querying social media site APIs. The implicit features include: age, personality, location, profile image, political bias. Due to space limitation, we ignore the description of age and personality. For explicit features, we have similar observations for explicit profile features as in [6], so we omit the discussion due to the space limitation. All features analysis are included here[6].

Location: Research has shown an inseparable relationship between user profiles and geo-locations. However, the location fields are usually very sparse. Thus, we exploit user-posted content to predict the user's location [7]. The idea is to identify "location indicative words" (LIW), which can encode an association with a particular location. The implementation of a pre-trained LIW model is integrated into an open source tool named pigeo [7], which is utilized here to predict the geo-locations of users in $\mathcal{U}^{(f)}$ and $\mathcal{U}^{(r)}$. The predicted results of pigeo are at the city-level and also include $(latitude, longitude)$ pairs and we observe that: (1) there are overall more users located in the US than other places, which is because most of the real/fake news items in our particular datasets are published and related to US politics and entertainments; and (2) the location distribution is different for fake and real news on both datasets, and the red and blue dots demonstrate the degree of differences. For example, there are general more real news share in east region of US in our datasets.

Profile Image: Profile images are important visual components of users on social media. Various studies have demonstrated the correlation between the choice of profile images with user personalities, behaviors, and activities. We classify the object types in profile images. With the recent development of deep learning in the computer vision domain, convolutional neural networks (CNN) have shown good performance for detecting objects in images. We chose the pre-trained VGG16 model [8] as it is the widely-used CNN architecture. We see that: the distributions of profile image classes are different for users in $\mathcal{U}^{(f)}$ and $\mathcal{U}^{(r)}$ on both datasets. For example, there are specific image types[7], such as "wig" and "mask" dominating the image categories for users spreading fake news, and "website" and "envelope" dominating the image categories for users spreading real news, on both datasets consistently.

Political Bias: Political bias plays an important role in shaping users' profiles and affecting their news consumption choices. Sociological studies on journalism demonstrate the correlation between partisan bias and news content authenticity (i.e., fake or real news) [9]. Reports have shown people's political affiliation is correlated with their attributes and behaviors[8]. Thus, we adopt method in [10] to measure user political bias scores by exploiting users' interests. The basic idea is that users who are more left-leaning or right-leaning have similar

[6]The omitted figures and analysis are available at https://tinyurl.com/y5mmdj2u

[7]http://image-net.org/explore

[8]https://2012election.procon.org/view.resource.php?resourceID=004818

interests among each other. We observe that: (1) users that are more likely to share fake news (i.e., $u \in \mathcal{U}^{(f)}$) also have a high probability to be biased on both datasets, and are more likely to be right-leaning; (2) users that are more likely to share real news (i.e., $u \in \mathcal{U}^{(r)}$) tend to be neutral-biased; and (3) overall, users in the two datasets demonstrate different political bias score distributions, indicating that the political bias of users could potentially help differentiate fake/real news.

In summary, we conclude that users in $\mathcal{U}^{(f)}$ and $\mathcal{U}^{(r)}$ reveal different feature distributions in most explicit and implicit feature fields, answering **RQ2**. These observations have great potential to guide the fake news detection process, which will be explored in detail in the next section.

IV. Exploiting User Profiles

In this section, we address **RQ3**. We explore whether the user profile features can help improve fake news detection, and how we can build effective models based on them, with feature importance and model robustness analysis.

A. Fake News Detection Performance

We first introduce how to extract user profile features $\mathbf{f}$ for news a. Let $\mathcal{U}$ denote the set of users who share news a. For each user $u_i \in \mathcal{U}$, we extract all types of aforementioned profile features and concatenate them into one feature vector $\mathbf{u}_i$. Note that for profile image features, since it has 1000 types, we use Principle Component Analysis to reduce the dimension to 10. Then we represent the user profile feature of a news as the average feature scores of all the users that share the news, i.e., $\mathbf{f} = \frac{1}{|\mathcal{U}|}\sum_{u_i \in \mathcal{U}} \mathbf{u}_i$. We also denote the proposed *U*ser *P*rofile *F*eature vector $\mathbf{f}$ as UPF.

To evaluate the performance of fake news detection algorithms, we use the following commonly used metrics to evaluate classifiers: Accuracy, Precision, Recall, and F1. We randomly choose 80% of news pieces for training and remaining 20% for testing, and the process is performed for 5 times and the average performance is reported. We compare the UPF with several state-of-the-art feature representations for fake news detection as below[9]:

- **RST** [11]: RST can capture the writing style of a document by extracting the rhetorical relations systematically. It learns a transformation from a bag-of-words surface representation into a latent feature representation[10].
- **LIWC** [12]: LIWC extracts lexicons that fall into different psycholinguistic categories, and learn a feature vector through multiple measures for each document[11].
- **RST_UPF**. RST_UPF represents the concatenated features of RST and UPF, which includes features extracted from both news content and user profiles.
- **LIWC_UPF**. LIWC_UPF represents the concatenated features of LIWC and UPF, which includes features extracted from both news content and user profiles.

We have the following observations:

- For news-content-based methods, we see that LIWC performs better than RST. This indicates that the LIWC vocabulary can better capture the deceptiveness in news content, which reveals that fake news pieces are very different from real news in terms of word choice from psychometrics perspectives.
- The UPF can achieve good performance in both datasets on all metrics. This shows that users that share more fake news and real news have different demographics and characteristics on social media, which serve as good features for fake news detection.
- In addition, RST_UPF performs better than either RST or UPF, which reveals that they are extracted from orthogonal information spaces, i.e., RST features are extracted from news content and UPF features from user profiles on social media, and have complementary information to help fake news detection.

TABLE II: Performance comparison for fake news detection with different feature representations.

	Metric	RST	LIWC	UPF	RST_UPF	LIWC_UPF
P	Acc	0.782	0.830	0.909	0.918	0.921
	Prec	0.777	0.809	0.948	0.949	0.942
	Recall	0.786	0.861	0.864	0.883	0.897
	F1	0.781	0.834	0.904	0.915	0.919
G	Acc	0.598	0.751	0.966	0.966	0.963
	Prec	0.601	0.796	0.956	0.952	0.949
	Recall	0.585	0.674	0.976	0.978	0.978
	F1	0.593	0.730	0.966	0.967	0.963

B. Feature Importance Analysis

Now we analyze the relative importance of these features for predicting fake news. We analyze feature importance in the Random Forest (RF) by computing a feature importance score based on the Gini impurity[12]. The top 5 common important features (with Gini impurity scores) are:

1) *RegisterTime* (0.937): the feature vector indicating the average distribution of verified and unverified users;
2) *Verified* (0.099): the feature vector indicating the average distribution of verified and unverified users;
3) *Political Bias* (0.063): the average bias score;
4) *Personality* (0.036): the average distribution of users' personality scores characterized by five factors distribution;
5) *StatusCount* (0.035): the average count of user posts.

We observe that (1) RegisterTime is the most important feature because newly created account may be more likely for fake news propagation purpose; (2) the distribution of verified/unverified user counts is important as verified users are less likely to spread fake news (3) the average political bias score is important because those users who share fake news are more likely to be biased to a specific ideology, while users that share real news tend to be least biased; (4) personality features are discriminative for detecting fake news because users' personalities affect their cognition and the way they respond to the real world [13]; and (5) the high importance score of StatusCount shows that the degrees of user activeness are quite different among users spreading fake and real news.

We further categorize the user profile features into three groups: **Explicit**, **Implicit** and **All** (i.e., both explicit and implicit features) and compare their contributions to the fake news detection task. The results are shown as in Table III. We observe that; (1) when all profile features are considered, the performance is higher than when only explicit or implicit features are considered. For example, the F1 scores on All features show a 4.51% and 9.84% increase compared with explicit

[9]All data and code are available at https://tinyurl.com/y5mmdj2u

[10]The code is available at: https://github.com/jiyfeng/DPLP

[11]The software and description of measures are available at: http://liwc.wpengine.com/

[12]http://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html. A higher Gini impurity score indicates a higher importance

TABLE III: Detection Performance with Different Group of Features from UPF.

	Feature Group	Acc	Prec	Recall	F1
P	All	**0.909**	**0.948**	**0.864**	**0.904**
	Explicit	0.870	0.891	0.841	0.865
	Implicit	0.837	0.892	0.763	0.823
G	All	**0.966**	**0.956**	**0.976**	**0.966**
	Explicit	0.894	0.884	0.906	0.895
	Implicit	0.961	0.956	0.967	0.962

and implicit feature groups on PolitiFact. This demonstrates that explicit and implicit features contain complementary information that can improve detection performance. (2) The implicit feature group is much more effective than the explicit feature group on Gossipcop for Accuracy and F1 scores. Note that implicit features require user-generated content to infer their values, which requires more effort to construct, while explicit features are often directly available in users' raw data. These observations allow us to better balance the trade-off with limited time and resources to make more informed decisions when building these feature groups.

V. Related Work

We briefly discuss work from (1) fake news detection on social media; and (2) measuring user profiles on social media.

A. Fake News Detection on Social Media

Fake news detection approaches generally fall into two categories depending on whether they use (1) *news content*; and (2) *social contexts* [1], [2]. For news content based approaches, features are extracted as linguistic-based such as writing styles [14], and visual-based such as fake images [15]. Social context based approaches incorporate features from social media user profiles, post contents, and social networks. User features measure users' characteristics and credibility. Post features represent users' social responses, such as stances [16]. Network features are extracted by constructing specific social networks, such as diffusion networks or co-occurrence networks. All of these social context models can basically be grouped as either stance-based or propagation-based. Stance-based models utilize users' opinions towards the news to infer news veracity [16]. Propagation-based models apply propagation methods to model unique patterns of information spread. To improve the explanatory power of fake new detection and to understand how to exploit user profiles to detect fake news, we perform, to our best knowledge, the first in-depth investigation of user profiles for their usefulness for fake news detection.

B. Measuring User Profiles on Social Media

User profiles on social media generally contain both *explicit* and *implicit* features. Explicit profile features (e.g., post count), which are already provided in raw user meta data, are widely exploited in different tasks on social media. While implicit profile features (e.g., personality), which are not directly provided, have proven very useful to apply to several specific analysis tasks. For age prediction, previous studies extract features from text posted by users. Schwartz *et al.* predicts gender, personality, and/or age simultaneously with open-vocabulary approaches [17]. For political bias prediction, existing works rely on tweets and hashtags, network structure, and language usage styles. Location prediction can be performed using "Location Indicative Words" (LIW) [7]. Profile images can be predicted using a pre-trained model in an unsupervised manner. We consider and extract both provided explicit and inferred implicit user profile features to better capture the different demographics of users for fake news detection.

VI. Conclusion and Future Work

In this work, we aim to answer questions regarding nature and extent of the correlation between user profiles on social media and fake news and provide a solution to utilize user profiles to detect fake news. This work opens up the doors for many areas of research. First, we will investigate the potential and foundation of other types of user feature in a similar way, such as content features and social network features, for fake news detection. Second, we will further investigate the correlations between malicious accounts and fake news to perform jointly detecting malicious accounts and fake news pieces. Third, we will explore various user engagement behaviors such as reposts, likes, comments, to further understand their utilities for fake news detection.

Acknowledgments

This material is in part supported by the ONR grant N000141812108, and Syracuse University CUSE grant.

References

[1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *KDD exploration newsletter*, 2017.
[2] X. Zhou and R. Zafarani, "Fake news: A survey of research, detection methods, and opportunities," *arXiv preprint arXiv:1812.00315*, 2018.
[3] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 675–684.
[4] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media," *arXiv preprint arXiv:1809.01286*, 2018.
[5] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, "Botornot: A system to evaluate social bots," in *WWW*, 2016, pp. 273–274.
[6] K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," in *MIPR*. IEEE, 2018.
[7] A. Rahimi, T. Cohn, and T. Baldwin, "pigeo: A python geotagging tool," *Proceedings of ACL-2016 System Demonstrations*, pp. 127–132, 2016.
[8] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Advances in neural information processing systems*, 2015, pp. 3483–3491.
[9] M. Gentzkow, J. M. Shapiro, and D. F. Stone, "Media bias in the marketplace: Theory," National Bureau of Economic Research, Tech. Rep., 2014.
[10] J. Kulshrestha, M. Eslami, J. Messias, M. B. Zafar, S. Ghosh, K. P. Gummadi, and K. Karahalios, "Quantifying search bias: Investigating sources of bias for political searches in social media," in *CSCW'17*.
[11] Y. Ji and J. Eisenstein, "Representation learning for text-level discourse parsing," in *ACL'2014*, vol. 1, 2014, pp. 13–24.
[12] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of liwc2015," Tech. Rep., 2015.
[13] N. Cantor and J. F. Kihlstrom, *Personality, cognition and social interaction*. Routledge, 2017, vol. 5.
[14] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," *arXiv preprint arXiv:1702.05638*, 2017.
[15] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, "Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy," in *WWW'13*.
[16] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News verification by exploiting conflicting social viewpoints in microblogs." in *AAAI*, 2016, pp. 2972–2978.
[17] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman *et al.*, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PloS one*, vol. 8, no. 9, p. e73791, 2013.