2018 IEEE Conference on Multimedia Information Processing and Retrieval

Understanding User Profiles on Social Media for Fake News Detection

Kai Shu
Arizona State University
Tempe, AZ 85281
kai.shu@asu.edu

Suhang Wang
Arizona State University
Tempe, AZ 85281
suhang.wang@asu.edu

Huan Liu
Arizona State University
Tempe, AZ 85281
huan.liu@asu.edu

***Abstract*—Consuming news from social media is becoming increasingly popular nowadays. Social media brings benefits to users due to the inherent nature of fast dissemination, cheap cost, and easy access. However, the quality of news is considered lower than traditional news outlets, resulting in large amounts of fake news. Detecting fake news becomes very important and is attracting increasing attention due to the detrimental effects on individuals and the society. The performance of detecting fake news only from content is generally not satisfactory, and it is suggested to incorporate user social engagements as auxiliary information to improve fake news detection. Thus it necessitates an in-depth understanding of the correlation between user profiles on social media and fake news. In this paper, we construct real-world datasets measuring users trust level on fake news and select representative groups of both "experienced" users who are able to recognize fake news items as false and "naïve" users who are more likely to believe fake news. We perform a comparative analysis over explicit and implicit profile features between these user groups, which reveals their potential to differentiate fake news. The findings of this paper lay the foundation for future automatic fake news detection research.**

***Keywords*-Fake News; User Profile; Trust Analysis;**

I. Introduction

Due to the increasing amount of our time spent online, people tend to seek out and receive news from social media. 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 seeing news on social media[1]. The reasons for the fast increase of users' engagement in news online are mainly because of the nature of social media such as the easy access, less expensive, and fast dissemination. Despite these advantages, the quality of news on social media is considered lower than that of traditional news outlets. Large amounts of fake news, i.e., those low quality news with intentionally false information [1], are widely spread online. For example, a report estimated that over 1 million tweets are related to fake news "Pizzagate"[2] by the end of 2016 presidential election.

Fake news has significant detrimental effects on individuals and the society. First, fake news intentionally misleads people to believe false information [2]. Second, fake news change the way people respond to real news. For example, people are confused about the news they read, which impedes their abilities to differentiating the truth from falsehood. Third, the trustworthiness of entire news ecosystem is broken due to fake news. Thus, it's critical to detect fake news on social media to mitigate these negative effects and to benefit the public and the news ecosystem.

However, detecting fake news on social media is quite challenging because it is written to intentionally mislead users, and attempt to distort truth with different styles while mimicking real news. Therefore, it's generally not satisfactory to detect fake news only from news content, and auxiliary information is needed, such as user engagements on social media. Recent research advancements make efforts to exploit user profiles which simply extract features without deep understanding of them, in which these features are like a black-box. Therefore, in this paper, we study the challenging problem of understanding user profiles on social media for fake news, which lays the foundation of using user profiles for fake news detection. In an attempt to understand the correlation between user profiles and fake news, we investigate the following research questions:

RQ1 *Are some users more likely to trust/distrust fake news?*
RQ2 *If yes to RQ1, what are the characteristics of these users that are more likely to trust/distrust fake news, and are there clear differences?*

By investigating **RQ1**, we try to understand if user profiles can be used for fake news detection. By answering **RQ2**, we can further provide guidance on which features of user profiles are useful for fake news detection. To answer these two questions, we conduct extensive statistical analysis on two real-world datasets. Our contributions are as follows:

- We study a challenging problem of investigating if user profiles can be used for fake news detection, which lays the foundation for improving fake news detection with user profiles;
- We propose a principled way to understand which features of user profiles are helpful for fake news detection, which ease the user profile feature construction for fake news detection; and
- We provide and experiment with two real-world datasets for studying the discriminative capacity of user profiles features.

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

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

0-7695-6354-6/18/$31.00 ©2018 IEEE
DOI 10.1109/MIPR.2018.00092

II. Related Work

In this section, we discuss the related work from two aspects. We first introduce some recent work for fake news detection on social media. Then, we discuss different ways on measuring user profiles on social media.

Fake News Detection on Social Media According to the sources that features are extracted from, fake news detection methods generally focus on using *news contents* and *social contexts* [1]. News content based approaches extract features from linguistic and visual information. Linguistic features aim to capture specific writing styles and sensational headlines that commonly occur in fake news content, such as lexical features and syntactic features [3]. Visual features identify fake images that are intentionally created or capture specific characteristics of images in fake news. Social context based approaches incorporate features from user profiles, post contents and social networks. User-based features measure users' characteristics and credibility [4]. Post-based features represent users' social responses such as stances [5], topics [6]. Network-based features are extracted by constructing specific networks, such as diffusion network [7], co-occurrence network [8], and propagation models can be further applied over these features [5]. Recently, Shu *et al.* propose to exploit the relationships among publisher, news, and users engagements on social media to detect fake news [9].

Existing approaches exploiting user profiles simply extract features to train classifiers without deep understanding of these features, which makes it a black-box. We perform an in-depth investigation on various aspects of user profiles for their usefulness for fake news detection.

Measuring User Profiles on Social Media User profiles can be generally categorized as *explicit* and *implicit* features. Explicit profile features, which are already provided in raw user meta data, are widely exploited in different tasks on social media, such as information credibility classification [4], user identity linkage [10]. While implicit profile features, which are not directly provided, are usually very useful to depict user profiles for specific tasks. Common implicit user features include age, gender, personality, etc. For gender prediction, Burger *et al.* exploit various text-based features from common explicit profile attributes and build SVM to build gender predictor [11]; Liu *et al.* incorporate user names as features and the method achieves a significant performance increase [12]. Personality inference on social media is first studied in [13] by using linguistic, structure and LIWC features and to predict the Big Five Personality Inventory [14] that shaping users' personalities. Celli proposes an unsupervised personality prediction model using various linguistic features [15]. For age prediction, previous studies focus on extracting features from text posted by users [16]. Schwartz *et al.* predict gender, personality and/or age simultaneously with open-vocabulary approaches [17].

Table I
The Statistics of Datasets

Platform	BuzzFeed	PolitiFact
# True news	91	120
# Fake news	91	120
# Users	15,257	23,865
# Engagements	25,240	37,259

We consider and extract both provided explicit and inferred implicit user profile features, to better capture the different demographics of users.

III. Assessing Users' Trust Level in News

In this section, we investigate **RQ1** by measuring the trust degree of users on social media towards fake and real news. First, we introduce the datasets that contain news and corresponding user engagements. Next, we identify user communities based on their trust levels in news.

A. Datasets

We construct two datasets with news content and social context information [3]. News content includes the meta attributes of the news, and social context includes the related user social engagements of news items. Determining the ground truth for fake/real news is challenging, usually requiring annotations with domain expertise with careful analysis. Existing datasets contain abundant news content information, but the ground truth information of fake/real news is very few. We use the ground truth labels collected from journalist experts from BuzzFeed[4] and well-recognized fact-checking website PolitiFact[5]. We bridge the gap between news content and social context by enriching the corresponding social engagements on social media. We generate the query using the news headlines and search them in Twitter API. Following traditional setting, we sample an equal number of real news to construct a balanced dataset, so that we can avoid trivial approach for fake news detection [5]. The details are shown in Table I.

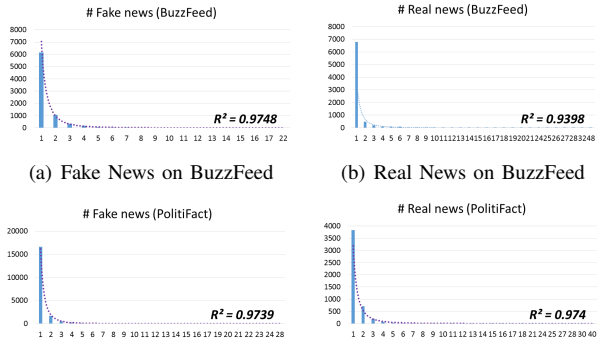
B. Identifying User Groups

We continue to investigate **RQ1** and identify different subset of users that reveal their trust degree in fake and real news. By find these groups, we want to build representative user sets that are more likely to trust fake/real news, from which we can further compare the degree of the differences of their profiles to find useful profile features. To measure the trust of user u_i, we analyze the user-news interactions, and compute the number of fake (real) news that user u_i has shared, denoted as $n_i^{(f)}$ ($n_i^{(r)}$). Intuitively, we assume

[3]https://github.com/KaiDMML/FakeNewsNet
[4]https://github.com/BuzzFeedNews/2016-10-facebook-fact-check.
[5]http://www.politifact.com/subjects/fake-news/

that if users share more fake news among all the users, they have higher trust degree of fake news; if users share more real news among all users, they have higher trust degree to real news. As shown in Figure 1, we plot the user count distribution with respect to the number of news they spread. We have the following observations: 1) In general, the majority of users only spread very few news pieces; 2) Few users spread many fake or real news pieces; 3) The distribution can fit in a power distribution with very high R^2 scores (≥ 0.93) in all cases, which shows that the distribution generally satisfy power-law distribution [18] with high significance values.

To choose those users that are more likely to trust fake (real) news, we select those "long tail" users who share the most absolute number of fake (real) news according to the distribution, and obtain two set of users $\mathcal{U}^{(f)} \subset \mathcal{U}$ and $\mathcal{U}^{(r)} \subset \mathcal{U}$, where $\mathcal{U}$ is the set of all users. We compute $\mathcal{U}^{(r)} = TopK(n_i^{(r)})$, indicating the top-$K$ users that share the most real news pieces; and $\mathcal{U}^{(f)} = TopK(n_i^{(f)})$, indicating the top-$K$ users that share the most fake news pieces.



(a) Fake News on BuzzFeed (b) Real News on BuzzFeed

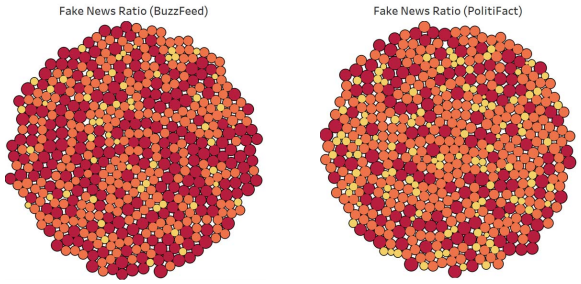
(c) Fake News on PolitiFact (d) Real News on PolitiFact

Figure 1. User Engaging Count Distribution

In addition, as another scenario, even though some users may not share many fake news items among all users, he/she may still share much more fake news than real news in his/her own history. To this end, we first divide all users into subsets that contains users i) only spread fake news; ii) only spread real news; and iii) spread both fake and real news, as shown in Table II. Then we propose to use another metric, named Fake News Ratio (FNR), defined as $FNR(i) = \frac{n_i^{(f)}}{n_i^{(r)}+n_i^{(f)}}$, where $FNR(i)$ denotes the FNR score of user u_i. The bigger the value, the larger the percentage that fake news items are being shared by u_i. We focus on those users that share both fake and real news (see Table II), and rank the FNR scores as shown in Figure 2. We set a threshold as 0.5 and if $FNR(i) > 0.5$ (red points in the figure), it indicates u_i shares more fake news than real news; if $FNR(i) < 0.5$ (yellow points), it indicates u_i shares more real news than fake news. Thus, we can enrich $\mathcal{U}^{(f)}$ by adding those users that satisfy $FNR(i) < 0.5$, enrich $\mathcal{U}^{(r)}$ by adding those users having $FNR(i) > 0.5$.

Table II
THE STATISTICS OF USER COMMUNITY

	BuzzFeed	PolitiFact
Only Fake	7,406	18,899
Only Real	7,316	4,437
Fake & Real	535	529
Total	15,257	23,865



(a) BuzzFeed (b) PolitiFact

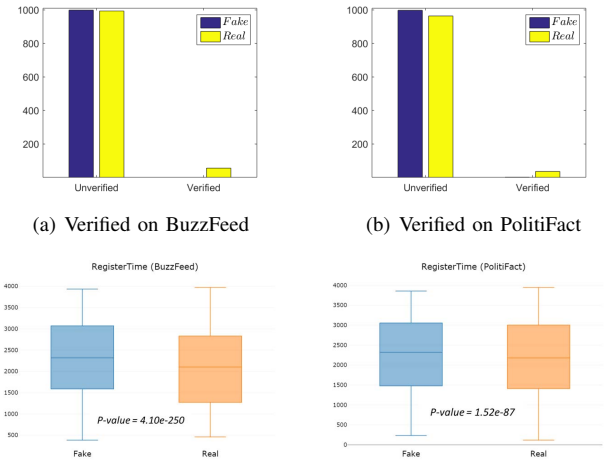
Figure 2. Fake News Ratio (FNR) Distribution.

Considering both the absolute number and FNR, we select Top-K users from those users in "Only Fake" or "Only Real" category by ranking the absolute number, and select additional users that in "Fake & Real" category through FNR scores. Finally, we empirically select $|\mathcal{U}^{(r)}| = |\mathcal{U}^{(f)}| = 1000$ users on BuzzFeed and PolitiFact datasets and $\mathcal{U}^{(r)} \cap \mathcal{U}^{(f)} = \phi$. We have now obtained users that are more likely to trust fake news $\mathcal{U}^{(f)}$ and real news $\mathcal{U}^{(r)}$, answering question **RQ1**. Note that we select two subsets $\mathcal{U}^{(f)}$ and $\mathcal{U}^{(r)}$ to ease the analytical process to find discriminate features in next section. The remaining users are ambiguous users that do not reveal clear trust preference on fake/real news in our datasets, which are excluded for the feature discriminate capacity analysis.

IV. CHARACTERIZING USER PROFILES

We select $\mathcal{U}^{(f)}$ and $\mathcal{U}^{(r)}$ because users in $\mathcal{U}^{(f)}$ are more likely to trust and share fake news, and those in $\mathcal{U}^{(r)}$ are more likely to share real news. However, to what extent and aspect these users are different is unknown. Thus, we continue to address **RQ2**, which involves measuring if there are clear differences among $\mathcal{U}^{(f)}$ and $\mathcal{U}^{(r)}$.

To this end, we collect and analyze user profile features from different aspects, i.e., *explicit* and *implicit*. Explicit features are obtained directly from metadata returned by querying social media site API. While implicit features are not directly available but inferred from user meta information or online behaviors, such as historical tweets. Our selected feature sets are by no means the comprehensive list



(a) Verified on BuzzFeed (b) Verified on PolitiFact

(c) RegisterTime on BuzzFeed (d) RegisterTime on PolitiFact

Figure 3. Profile Feature Comparison.

of all possible features. However, we focus on those explicit features that can be easily accessed and almost available for all public users, and implicit features that are widely used to depict user profiles for better understanding user characteristics for guiding informative features.

A. Explicit Profile Features

We first analyze those profile fields in the meta information that describe a user on social media. A list of representative attributes include:

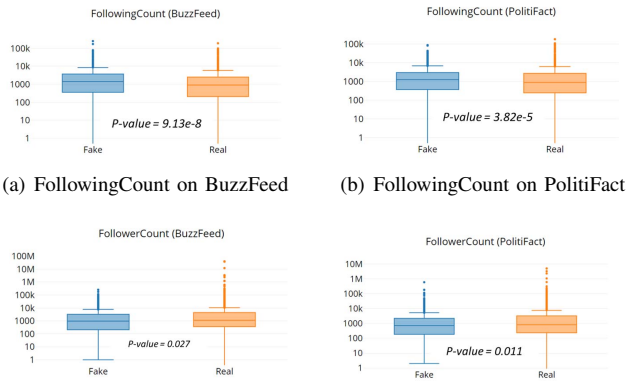
- **Profile-Related**: The basic user description fields.
  - *Verified*: indicating whether it is a verified user;
  - *RegisterTime*: the number of days past since the accounted is registered;
- **Content-Related**: The attributes of user activities.
  - *StatusCount*: The number of posts;
  - *FavorCount*: The number of favorite action the user performs.
- **Network-Related**: The social networks attributes;
  - *FollowerCount*: The number of follower count;
  - *FollowingCount*: The number of following count;

Next, we compare these fields to demonstrate whether the users in $\mathcal{U}^{(r)}$ and $\mathcal{U}^{(f)}$ have clear differences. If a feature f reveal clear differences between $\mathcal{U}^{(r)}$ and $\mathcal{U}^{(f)}$, then f has the potential usefulness for detecting fake news; otherwise, f may not be useful for fake news detection.

Profile-related features are compared in Figure 3. We rank RegisterTime values of all users in $\mathcal{U}^{(f)}$ and $\mathcal{U}^{(r)}$ and perform two-tail statistical t-test with significant level 0.05 on the corresponding ranking pairs. If the p value is less than significance level (i.e., 0.05), then it exhibit significant difference between $\mathcal{U}^{(f)}$ and $\mathcal{U}^{(r)}$. We can see that there are more verified users in $\mathcal{U}^{(r)}$ than $\mathcal{U}^{(f)}$ on BuzzFeed and PolitiFact significantly, which shows that verified users are more likely to trust real news. In addition, the box-and-whisker diagram shows the distribution of user register time exhibit a significant difference between both user groups. The observations on both datasets demonstrate that users registered earlier are more likely to trust fake news, and newer accounts tend to spread more real news, which are consistent with previous work [4].

We compare content-related profile features in Figure 5. Similarly, we rank the StatusCount and FavorCount and perform t-test, and we have the following observations on both datasets; 1) The users in $\mathcal{U}^{(f)}$ generally publish fewer posts that users in $\mathcal{U}^{(r)}$, which indicates those users trusting more real news are more likely to be active and express themselves; 2) The users in $\mathcal{U}^{(f)}$ tends to express more "favor" actions to tweets posted by other users, indicating their willingness to reach out to other users.

We compare network-related profile features as in Figure 4 and Figure 6. From Figure 4, we see that user in $\mathcal{U}^{(f)}$ have fewer followers and more following counts (as absolute values) from various metrics such as Median, MAX and Mean values on both datasets. To further measure the relative number of following and followers, we also compute the TFF Ratio, indicating the ratio of follower to following counts [6]. The distribution of TFF ratio values is in Figure 6, which shows that: 1) $\mathcal{U}^{(f)}$ includes more users that with $TFF < 1$, indicating less followers than followings; 2) $\mathcal{U}^{(r)}$ consistently has more users with $TFF > 1$, indicating users trust real news are more likely to be more popular.



(a) FollowingCount on BuzzFeed (b) FollowingCount on PolitiFact

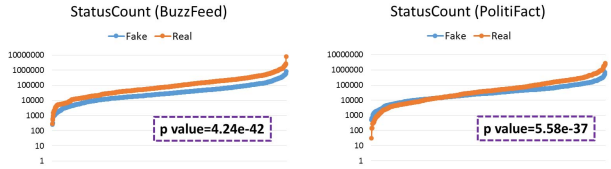
(c) FollowerCount on BuzzFeed (d) FollowerCount on PolitiFact

Figure 4. Network Feature Comparison.

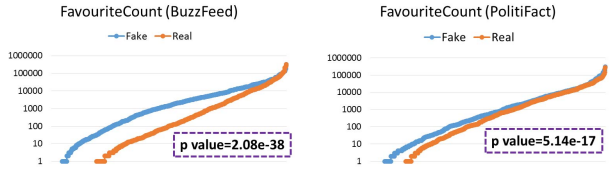
B. Implicit Profile Features

To further investigate the characteristics in $\mathcal{U}^{(f)}$ and $\mathcal{U}^{(r)}$, we explore several implicit profile features: gender, age, and personality. We choose these three features because they are not directly provided with user meta data but they are commonly used to better describe users [17].

[6]We use adapted TFF computed as $TFF = \frac{\#Follower+1}{\#Following+1}$.

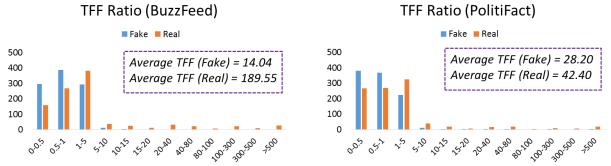


(a) StatusCount on BuzzFeed (b) StatusCount on PolitiFact



(c) FavorCount on BuzzFeed (d) FavorCount on PolitiFact

Figure 5. Content Feature Comparison.



(a) BuzzFeed (b) PolitiFact

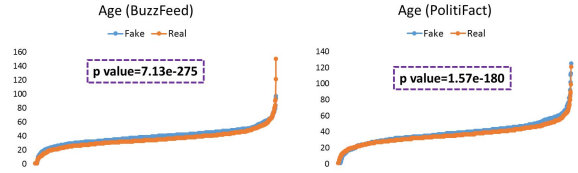
Figure 6. TFF Ratio Score Comparison.

Gender and Age. Studies have shown that gender and age have major impacts on people's psychology and cognition. For example, men and women generally differ markedly in their interests and work preferences [19]. With age gradually change, people typically become less open to experiences but more agreeable and conscientious [20]. We aim to answer the questions: 1) Are female or male users more likely to trust fake news; 2) Are users in different ages have different abilities to differentiate fake news?

To this end, we infer the gender and age of users using existing state-of-the-art approach [21]. The idea is to build a linear regression model with the collected predictive lexica (with words and weights). We utilize the recent post Tweets as the corpus to extract relevant words in the lexica. The results are show as in Figure 7 and 8:

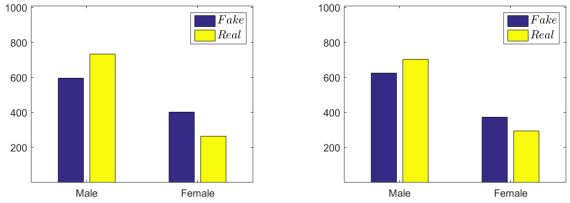
- Generally, the predicted ages are slightly bigger for users in $\mathcal{U}^{(f)}$ than those in $\mathcal{U}^{(r)}$. It shows older people are more likely to trust fake news.
- More male users than female users on social media are engaging in news consumption. In addition, the ratio of male to female is higher among those users in $\mathcal{U}^{(r)}$, which indicates female users are more likely to trust fake news than male users.

Personality. Personality refers to the traits and characteristics that make an individual different from others. Following traditional setting, we draw on the popular Five Factor Model (or "Big Five"), which classifies personality



(a) BuzzFeed (b) PolitiFact

Figure 7. Age Distribution Comparison.

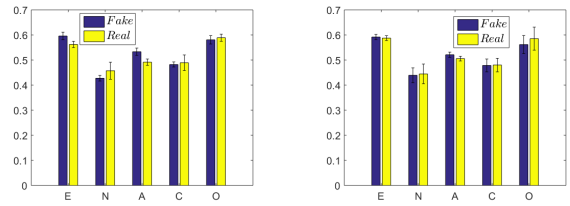


(a) BuzzFeed (b) PolitiFact

Figure 8. Gender Distribution Comparison.

traits into five dimensions: **E**xtraversion (e.g., outgoing, talkative, active), **A**greeableness (e.g., trusting, kind, generous), **C**onscientiousness (e.g., self-controlled, responsible, thorough), **N**euroticism (e.g., anxious, depressive, touchy), and **O**penness (e.g., intellectual, artistic, insightful). We try to answer the following question: do personalities clearly exist between users that are more likely to trust fake news and those more likely to trust real news?

To predict users' personalities while no ground truth available, we apply an unsupervised personality prediction tool named Pear [15], a pre-trained model using the recent tweets posted by users[7]. It is one of the state-of-the-art unsupervised text-based personality prediction model. As shown in Figure 9, we can see that on both datasets: 1) The users tend to have relatively high Extraversion and Openness , and low Neuroticism scores, indicating more outgoing and intellectual, and less calm for personality; 2) Users in $\mathcal{U}^{(r)}$ tends to have higher Extraversion and Agreeableness scores than those in $\mathcal{U}^{(f)}$, indicating that users are extrovert and friendly are more likely to trust real news.



(a) BuzzFeed (b) PolitiFact

Figure 9. Personality Scores Comparison.

In sum, we conclude that users in $\mathcal{U}^{(f)}$ and $\mathcal{U}^{(r)}$ reveal

[7]http://personality.altervista.org/pear.php

different feature distributions in most explicit and implicit feature fields, answering **RQ2**. These observations have great potential to guide the fake news detection process, and we would like to explore it in future work.

V. Conclusion and Future Work

Due to the potential of using user engagements on social media to help fake news detection, we investigate the correlation between user profiles and fake/real news. Experimental results on real-world datasets demonstrate that: i) there are specific users who are more likely to trust fake news than real news; and ii) these users reveal different features from those who are more likely to trust real news. These observations ease the feature construction of profiles features for fake news detection.

There are several interesting future directions. First, we want to explore other user profile features, such as political bias and user credibility, to better understand if these user characteristics can be used for fake news detection. Second, in this paper, we identify a set of potential user profiles for fake news detection. We want to further investigate how these features can be aggregated to fake news detection models to advance fake news detection. Third, research has shown that fake news pieces have been widely spread by bots, and we will incorporate bot detections techniques to discriminate bots from normal users for better exploiting user profile features to detect fake news.

Acknowledgments

We thank the reviewers for insightful feedbacks. This material is based upon work supported by, or in part by, the ONR grant N00014-16-1-2257.

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] B. Nyhan and J. Reifler, "When corrections fail: The persistence of political misperceptions," *Political Behavior*, vol. 32, no. 2, pp. 303–330, 2010.

[3] 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.

[4] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *WWW'11*.

[5] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News verification by exploiting conflicting social viewpoints in microblogs." in *AAAI*, 2016, pp. 2972–2978.

[6] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong, "Detect rumors using time series of social context information on microblogging websites," in *CIKM'15*.

[7] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," in *ICDM'13*. IEEE, 2013, pp. 1103–1108.

[8] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news," *arXiv preprint arXiv:1703.06959*, 2017.

[9] K. Shu, S. Wang, and H. Liu, "Exploiting tri-relationship for fake news detection," *arXiv preprint arXiv:1712.07709*, 2017.

[10] K. Shu, S. Wang, J. Tang, R. Zafarani, and H. Liu, "User identity linkage across online social networks: A review," *ACM SIGKDD Explorations Newsletter*, vol. 18, no. 2, pp. 5–17, 2017.

[11] J. D. Burger, J. Henderson, G. Kim, and G. Zarrella, "Discriminating gender on twitter," in *EMNLP'11*.

[12] W. Liu and D. Ruths, "What's in a name? using first names as features for gender inference in twitter." in *AAAI spring symposium: Analyzing microtext*, vol. 13, 2013, p. 01.

[13] J. Golbeck, C. Robles, and K. Turner, "Predicting personality with social media," in *CHI'11 extended abstracts on human factors in computing systems*. ACM, 2011, pp. 253–262.

[14] G. M. Hurtz and J. J. Donovan, "Personality and job performance: The big five revisited," *Journal of applied psychology*, vol. 85, no. 6, pp. 869–879, 2000.

[15] F. Celli and M. Poesio, "Pr2: A language independent unsupervised tool for personality recognition from text," *arXiv preprint arXiv:1402.2796*, 2014.

[16] D.-P. Nguyen, R. Gravel, R. B. Trieschnigg, and T. Meder, "" how old do you think i am?" a study of language and age in twitter," 2013.

[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.

[18] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[19] R. Su, J. Rounds, and P. I. Armstrong, "Men and things, women and people: a meta-analysis of sex differences in interests." 2009.

[20] R. R. McCrae, P. T. Costa, M. P. de Lima, A. Simões, F. Ostendorf, A. Angleitner, I. Marušić, D. Bratko, G. V. Caprara, C. Barbaranelli *et al.*, "Age differences in personality across the adult life span: parallels in five cultures." *Developmental psychology*, vol. 35, no. 2, p. 466, 1999.

[21] M. Sap, G. Park, J. Eichstaedt, M. Kern, D. Stillwell, M. Kosinski, L. Ungar, and H. A. Schwartz, "Developing age and gender predictive lexica over social media," in *EMNLP'14*.