Profiling fake news spreaders: stylometry, personality, emotions and embeddings
Notebook for PAN at CLEF 2020

Elisabetta Fersini, Justin Armanini, and Michael D'Intorni

University of Milano-Bicocca, Viale Sarca 336, Milan - Italy
elisabetta.fersini@unimib.it, j.armanini@campus.unimib.it,
m.dintorni@campus.unimib.it

Abstract This paper describes our proposed solution for the Profiling Fake News Spreaders on Twitter shared task at PAN 2020 [23]. The task consists in determining whether a given author a set of Twitter posts is a fake news spreader or not, both for the English and Spanish languages. The proposed approach is based on modeling both types of users according to four main types of characteristics, i.e. stylometry, personality, emotions and feed embeddings. Our system achieved an accuracy of 60% for the English dataset, while 72% for the Spanish one.

1 Introduction

The problem of fake news and rumour detection has gained a lot of attention during the last years. If users get their news from social media, they encounter the risk of being exposed to false or misleading content such as hoaxes, rumors and click-bait headlines. The main advantages for fake news providers is to generate traffic to specific web sites in order to monetize through advertising [2] or to manipulate politically related facts [24]. Since the beginning, the massive spread of fake news has been identified as a major global risk [32]. Countering of fake news is a challenging problem that can be addressed at two different levels: when fake contents are created and when fake contents are spread. Concerning the recognition of fake news, several approaches have been proposed into the state of the art, ranging from unsupervised [33] to semi-supervised [25] and supervised approaches [11]. On the other hand, to the best of our knowledge, there isn't any study for predicting if a given user is more inclined to share a fake or a real news in online social networks.

The proposed approach, presented for the "Profiling Fake News Spreaders on Twitter" challenge organized at the PAN@CLEF initiative, is one of the first tentatives for trying to prevent the intentional or unintentional diffusion of inaccurate information. In particular, we propose to characterize the profile of fake news and real news spreaders, by exploiting four types of characteristics, i.e. stylometry, personality, emotions and embeddings.

2 Related work

Fake news is a phenomenon that has started to grow exponentially during the last ten years [17]. The leading cause is that fake news can be created and shared online very quickly, in a not expensive way when compared to traditional news media such as newspapers. Most of the literature available in this research area is mostly related to the recognition of fake news. In particular, several approaches are nowadays available in the state of the art, ranging from methods based on stylistic features and patterns, to those focused on the source credibility and to the ones concentrated to the propagation dynamics.

Concerning the approaches grounded on stylistic features, the content is commonly represented by a set of characteristics [26,19,24,1,6,5,35,7,22], that are consequently used by any machine learning approach. Examples of such features are readability, psycholinguistic features, punctuation and syntax. For what concerns the methods focused on source credibility, most of them are based on evaluating the users that have created a potential false information by using several cues such as emotions [8,9], users posting and re-posting behavior [30,15] and content-specific features [10]. Regarding the approaches focused on the propagation dynamics of fake news, most of the them are based on epidemic models, which can mathematically model the progression of an infectious disease [31,16,3]. However, defining proper epidemic model for fake news is still in its infancy due to their assumptions that in many cases do not match a real scenario.

In all of the above mentioned approaches, the users play a key role in the creation and propagation of fake news by consuming and spreading contents that could be fake or real. To the best of our knowledge, few studies [27,28,29] focus on profiling possible fake news spreaders in online social media. In this paper, grounding on the main findings achieved by the former approaches, we aimed at modelling fake news spreaders by profiling users exploiting the most promising characteristics available in the state of the art. The proposed method, based on stylometry, personality, emotions and embeddings is detailed in the following section.

3 Proposed Method

The proposed system aims at distinguishing authors that have shared some fake news in the past from those that have never done it, by characterizing each user according to the following features:

- Stylometry: the writing style of a user can reveal if it is mainly inclined to fake or real contents. To this purpose, for each user, we estimated its stylometric profile by considering language usage, punctuation adoption and part-of-speech frequencies;
- Personality: the behavior of users in social media, such the posts that they create or the contents that they share, allows us to infer sensitive information such as personality. Taking into account the Twitter feeds, each user can be associated to a given personality according its communication skills;

- Emotion: since real news define their contents without attempting to affect the opinion of the reader and, on the other hand, fake news take advantage of the readers sensitivity, we modeled the emotion conveyed by each user through the spread text;
- Embeddings: since fake news spreaders should be inclined, intentionally or unintentionally, to share specific topics of interest, an embedding representation of such contents have been derived.

A summary of the proposed vector representation for each user is reported in Figure 1, where 8-features are used to describe the emotion of the user, 1-feature for its personality, 32-features for characterizing its writing style and 512-features for its content embeddings.

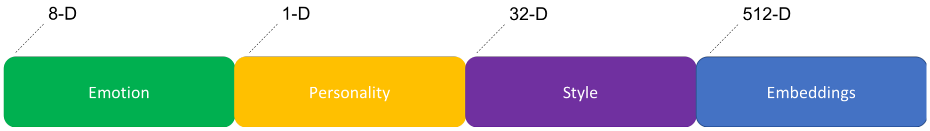

Figure 1. Proposed user representation

In the following sub-sections, the features extracted for modeling fake and real news spreaders will be detailed. Once all the above mentioned characteristics have been extracted for each user, a Support Vector Machines (SVM) classifier, with a linear kernel and default parameters according to the scikit-learn implementation [18] has been adopted for distinguishing between fake and real news spreaders. The proposed model has been run on a TIRA architecture [21].

3.1 Stylometric characteristics

Stylometric features can be used as baseline characteristics to profile fake news and real news spreaders, to therefore train a model to distinguish them. The stylometric features investigated in this paper are the following ones:

- *language usage*: average number of sentences, average number of words, average number of # symbols, @ symbol, frequency of unique words, frequency of complex words (more than 5 characters), average of the number of characters in a word, frequency of emoji, frequency of offensive words, frequency of stretched words, frequency of upper-case words, frequency of words starting with upper case, frequency of named-entities;
- *part-of-speech*: frequency of verbs, auxiliary verbs, adjectives, superlative adjectives, superlative relative adjectives, comparative adjectives, nouns, conjunctions, adverbs, articles, indefinite articles, pronouns, numbers, first singular person pronouns, first plural person pronouns, second singular person pronouns, second plural person pronouns, third singular person pronouns related to male, third singular person pronouns related to female, third plural person pronouns related to male, third plural person pronouns related to female;

- *punctuation marks*: frequency of punctuation, colon, semi-comma, exclamation mark, question mark, quotes,

3.2 Personality traits

In order to validate the hypothesis that some personality traits are more keen to spread fake news than others, we exploited the model based on the Myers Briggs Type Indicator (MBTI) [14] to predict the personality type of a given user. In particular, 16 distinct personality types are detected by a 4 axis model that compares the following dichotomies:

- Introversion (I) *vs* Extroversion (E)
- Intuition (N) *vs* Sensing (S)
- Thinking (T) *vs* Feeling (F)
- Judging (J) *vs* Perceiving (P)

The choice of the MTBI model has been motivated by the hypothesis that real news spreaders should belong more likely to the type T and J, for being more predisposed to reasoning and accurate decision making. On the contrary, we argue that fake news spreaders should have a personality type E or F, for being more inclined to act according to their feelings. In order to detect the personality type of each user, we adopted the MBTI personality classification system that takes the social media posts of a given user as input and produces as output a prediction of the author's personality type. To accomplish this task, only for the English language, we exploited a supervised model [1] based on a Naive Bayes classifiers trained using a publicly available Kaggle dataset[2]. The MTBI classifiers is based on two main components: (1) pre-processing of Tweeter feeds posted by the user and (2) training/inference mechanism based on Naive Bayes for predicting the personality of a user given its posts.

The pre-processing component, based on NLTK [12], lemmatizes the text, in order to transform the inflected forms of the same root word to their lemma. Then, through the Keras word tokenizer, 2500 most common lemmatized words of the lemmatized text are maintained for the subsequent steps. The n-grams and word vectors for the hashtags, emoticons and phrases are created by using the TF-IDF representation. Concerning the training and inference mechanisms, a Naive-Bayes Text Classifier is adopted for predicting the personality of a user give its Twitter feeds.

3.3 Emotion-related features

Extracting the emotion-related feature for each user is the first step for characterizing its profile. In order to extract emotions, the "NRC Emotion Lexicon" has been used [13], which consists of a word list with associated the eight emotions (anger, fear, expectation, trust, surprise, sadness, joy and disgust) modeled by the Plutchick theory [20]. The

[1] https://github.com/priyansh19/Classification-of-Personality-based-on-Users-Twitter-Data

[2] https://www.kaggle.com/datasnaek/mbti-type

lexicon is composed of 14182 words, and was created through "crowdsourcing" starting from an idea by Mohammad Saif and Peter Turney. Initially developed using only English words, in 2017 it was extended for supporting multiple languages. For creating the emotion-related features, the frequency of words belonging to the eight emotions have been estimated using all the Twitter feeds of a given user.

3.4 User Embeddings

The last characteristics that we included for representing our users are related to embeddings. The hypothesis is that there are some aspects of a similar fake news that could be expressed in a similar way from a semantic point of view even if they are written in different ways from a lexicographic perspective. In order to capture semantic similarities of fake news spreaders, each user has been represented by a 512-D vector derived by a member-wise mean aggregation function on its tweet embedding. To this purpose, we adopted the Universal Sentence Encoder (v4) [4] developed by Google and available in the TensorFlow Hub package. In particular, to capture some common characteristics between the two considered languages, the Multilingual Universal Sentence Encoder [34] has been adopted. This model is an extension of the Universal Sentence Encoder Large that includes training on multiple tasks across languages. The process for extracting the embedding representation of a given user is reported if Figure 2.

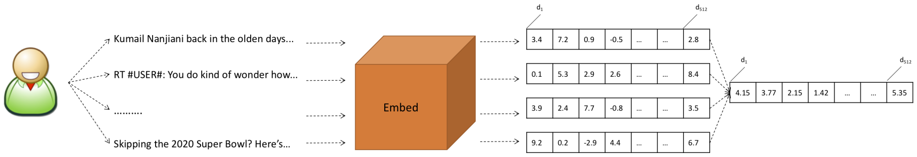

Figure 2. User embeddings

4 Experiments and Results

4.1 Experimental settings

For evaluating the performance of the systems aimed at distinguishing fake news spreaders from real news spreaders, the task organizers provided both training and testing datasets, for the English and Spanish language. The training datasets are composed of 600 users, perfectly balanced between fake and real news spreaders, with their corresponding 100 messages. Analogous datasets have been provided for the testing phase. The proposed method has been validated by measuring the accuracy using a 10-cross validation strategy on the training set, and on the testing set given for the competition.

4.2 Experimental results

We report in the following the results firstly obtained by adopting a 10-cross validation strategy on the training dataset. Table 1 shows the accuracy obtained on each fold, together with the average performance and its standard deviation, on the two languages. We can easily note that the results on the Spanish dataset are a bit higher than the ones obtained on the English language. This is likely due to the variability of topics considered in both languages.

	Fold											
	1	2	3	4	5	6	7	8	9	10	Avg.	Std.
English	0.70	0.60	0.53	0.80	0.60	0.67	0.63	0.57	0.57	0.57	0.62	0.08
Spanish	0.77	0.73	0.77	0.67	0.73	0.87	0.80	0.73	0.60	0.80	0.75	0.07

Table 1. 10-fold cross-validation results.

By analysing the performance of the proposed approach, we notice that the features related to stylometry, emotions and embeddings contribute more to the recognition capabilities than the personality one. This is due to the inefficacy of the adopted model to really capture the personality traits of the users given their Twitter feeds. Another interesting insights is related to the relative contribution given by the stylometric characteristics to the final results. These features contributed to obtain a 5% of improvement of the accuracy, with respect to use only emotion and embeddings. This reveals that this set of characteristics can help to better distinguish fake and not-fake writing styles.

Concerning the results of the shared task, the proposed approach achieved 60% of accuracy for English and 72% for Spanish.

4.3 Conclusions and future work

This paper has presented the proposed solution for the Profiling Fake News Spreaders on Twitter shared task at PAN 2020. Our approach, based on modeling both fake and real news spreaders of users using stylometry, personality, emotions and embeddings, has shown promising results and pointed out interesting research directions. Concerning the obtained results, the analysis of the considered characteristics has highlighted that stylometry could play an important role for characterising both profiles, while personality does not contribute in a significant way. Regarding future work, some additional characteristics should be considered. For instance, age, geo-location and education level of the users could contribute to better distinguish between the two profiles. Other research directions are focused on the analysis of the syntactic patters and relationships between sentences.

References

1. Biyani, P., Tsioutsiouliklis, K., Blackmer, J.: âĂIJ8 amazing secrets for getting more clicksâĂİ: Detecting clickbaits in news streams using article informality. In: Proceedings of

the Thirtieth AAAI Conference on Artificial Intelligence. p. 94âĂŞ100. AAAIâĂŹ16, AAAI Press (2016)

2. Braun, J.A., Eklund, J.L.: Fake news, real money: Ad tech platforms, profit-driven hoaxes, and the business of journalism. Digital Journalism **7**(1), 1–21 (2019)
3. Carchiolo, V., Longheu, A., Malgeri, M., Mangioni, G., Previti, M.: A trust-based news spreading model. In: International Workshop on Complex Networks. pp. 303–310. Springer (2018)
4. Cer, D., Yang, Y., Kong, S.y., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Strope, B., Kurzweil, R.: Universal sentence encoder for English. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. pp. 169–174. Association for Computational Linguistics, Brussels, Belgium (Nov 2018)
5. Feng, S., Banerjee, R., Choi, Y.: Syntactic stylometry for deception detection. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 171–175 (2012)
6. Ghanem, B., Rosso, P., Rangel, F.: An Emotional Analysis of False Information in Social Media and News Articles. ACM Transactions on Internet Technology (TOIT) **20**(2), 1–18 (2020)
7. Giachanou, A., Ríssola, E.A., Ghanem, B., Crestani, F., Rosso, P.: The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. In: International Conference on Applications of Natural Language to Information Systems. pp. 181–192. Springer (2020)
8. Giachanou, A., Rosso, P., Crestani, F.: Leveraging emotional signals for credibility detection. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 877âĂŞ880. SIGIRâĂŹ19, Association for Computing Machinery, New York, NY, USA (2019)
9. Guo, C., Cao, J., Zhang, X., Shu, K., Liu, H.: Dean: Learning dual emotion for fake news detection on social media. arXiv preprint arXiv:1903.01728 (2019)
10. Gupta, A., Kumaraguru, P., Castillo, C., Meier, P.: TweetCred: Real-Time Credibility Assessment of Content on Twitter, pp. 228–243. Springer International Publishing, Cham (2014)
11. Kaliyar, R.K., Goswami, A., Narang, P., Sinha, S.: Fndnet–a deep convolutional neural network for fake news detection. Cognitive Systems Research **61**, 32–44 (2020)
12. Loper, E., Bird, S.: Nltk: The natural language toolkit. In: Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics. pp. 63–70 (2002)
13. Mohammad, S.M., Turney, P.D.: Crowdsourcing a word–emotion association lexicon. Computational Intelligence **29**(3), 436–465 (2013)
14. Myers, I.B., Myers, P.B.: Gifts Differing: Understanding Personality Type. Hachette UK (2010)
15. ODonovan, J., Kang, B., Meyer, G., Höllerer, T., Adalii, S.: Credibility in context: An analysis of feature distributions in twitter. In: 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing. pp. 293–301. IEEE (2012)
16. Oktaviansyah, E., Rahman, A.: Predicting hoax spread in indonesia using sirs model. In: Journal of Physics: Conference Series. vol. 1490, p. 012059. IOP Publishing (2020)
17. Parikh, S.B., Patil, V., Atrey, P.K.: On the origin, proliferation and tone of fake news. In: 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). pp. 135–140. IEEE (2019)

18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. the Journal of machine Learning research **12**, 2825–2830 (2011)
19. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R.: Automatic detection of fake news. In: Proceedings of the 27th International Conference on Computational Linguistics. pp. 3391–3401 (2018)
20. Plutchik, R.: Emotions: A general psychoevolutionary theory. Approaches to emotion **1984**, 197–219 (1984)
21. Potthast, M., Gollub, T., Wiegmann, M., Stein, B.: TIRA Integrated Research Architecture. In: Ferro, N., Peters, C. (eds.) Information Retrieval Evaluation in a Changing World. Springer (Sep 2019)
22. Rangel, F., Franco-Salvador, M., Rosso, P.: A Low Dimensionality Representation for Language Variety Identification. In: International Conference on Intelligent Text Processing and Computational Linguistics. pp. 156–169. Springer (2016)
23. Rangel, F., Giachanou, A., Ghanem, B., Rosso, P.: Overview of the 8th Author Profiling Task at PAN 2020: Profiling Fake News Spreaders on Twitter. In: Cappellato, L., Eickhoff, C., Ferro, N., Névéol, A. (eds.) CLEF 2020 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (Sep 2020), CEUR-WS.org
24. Rashkin, H., Choi, E., Jang, J.Y., Volkova, S., Choi, Y.: Truth of varying shades: Analyzing language in fake news and political fact-checking. In: Proceedings of the 2017 conference on empirical methods in natural language processing. pp. 2931–2937 (2017)
25. Saini, N., Singhal, M., Tanwar, M., Meel, P.: Multimodal, semi-supervised and unsupervised web content credibility analysis frameworks. In: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). pp. 948–955. IEEE (2020)
26. Schuster, T., Schuster, R., Shah, D.J., Barzilay, R.: The limitations of stylometry for detecting machine-generated fake news. Computational Linguistics pp. 1–12 (2020)
27. Shu, K., Wang, S., Liu, H.: Understanding user profiles on social media for fake news detection. In: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). pp. 430–435. IEEE (2018)
28. Shu, K., Wang, S., Liu, H.: Beyond news contents: The role of social context for fake news detection. In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. p. 312âĂŞ320. WSDM âĂŹ19, Association for Computing Machinery, New York, NY, USA (2019)
29. Shu, K., Zhou, X., Wang, S., Zafarani, R., Liu, H.: The role of user profiles for fake news detection. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. p. 436âĂŞ439. ASONAM âĂŹ19, Association for Computing Machinery, New York, NY, USA (2019)
30. Sitaula, N., Mohan, C.K., Grygiel, J., Zhou, X., Zafarani, R.: Credibility-based fake news detection. In: Disinformation, Misinformation, and Fake News in Social Media, pp. 163–182. Springer (2020)
31. Tambuscio, M., Ruffo, G., Flammini, A., Menczer, F.: Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. In: Proceedings of the 24th international conference on World Wide Web. pp. 977–982 (2015)
32. Webb, H., Jirotka, M., Stahl, B.C., Housley, W., Edwards, A., Williams, M., Procter, R., Rana, O., Burnap, P.: 'digital wildfires' a challenge to the governance of social media? In: Proceedings of the ACM web science conference. pp. 1–2 (2015)
33. Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., Liu, H.: Unsupervised fake news detection on social media: A generative approach. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 5644–5651 (2019)

34. Yang, Y., Cer, D., Ahmad, A., Guo, M., Law, J., Constant, N., Abrego, G.H., Yuan, S., Tar, C., Sung, Y.H., et al.: Multilingual universal sentence encoder for semantic retrieval. arXiv preprint arXiv:1907.04307 (2019)
35. Zhou, X., Jain, A., Phoha, V.V., Zafarani, R.: Fake news early detection: A theory-driven model. Digital Threats: Research and Practice **1**(2), 1–25 (2020)