Assembly of polarity, emotion and user statistics for detection of fake profiles
Notebook for PAN at CLEF 2020

Luis Gabriel Moreno-Sandoval[1,3], Edwin Puertas[2,1,3], Alexandra
Pomares-Quimbaya[1,3], and Jorge Andres Alvarado-Valencia[1,3]

[1] Pontificia Universidad Javeriana, Bogotá, Colombia
`{morenoluis,edwin.puertas,pomares,jorge.alvarado}@javeriana.edu.co`
[2] Universidad Tecnológica de Bolívar, Cartagena, Colombia
`epuerta@utb.edu.co`
[3] Center of Excellence and Appropriation in Big Data and Data Analytics (CAOBA)

Abstract The explosive growth of fake news on social networks has aroused
great interest from researchers in different disciplines. To achieve efficient and
effective detection of fake news requires scientific contributions from various dis-
ciplines, such as computational linguistics, artificial intelligence, and sociology.
Here we illustrate how polarity, emotion, and user statistics can be used to detect
fake profiles on Twitter's social network. This paper presents a novel strategy for
the characterization of the Twitter profile based on the generation of an assem-
bly of polarity, emotion, and user statistics characteristics that serve as input to a
set of classifiers. The results are part of our participation in the PAN 2020 in the
CLEF in the task of Profiling Fake News Spreaders on Twitter.

1 Introduction

The exponential growth in social networks of fake news and rumors has led researchers
from different areas to join efforts to quickly and accurately mitigate these phenomena'
proliferation. Thus, the PAN at CLEF of the 2020 edition has proposed a task of au-
thorship analysis whose objective is to identify possible fake news spreaders [16] in the
social networks as a first step to avoid the propagation of the already fake news said
amid the online users.

The way we collect and consume news has become a crucial process these days
due to the growth of social media platforms, such as the social networks Twitter [4] and
Facebook [5], which have reported an exponential increase in popularity [3,18]. As an
example, Twitter reported 330 million active users per month in early 2020. [6] Mean-
while, Facebook reported 2.603 million active Facebook users per month worldwide as

[4] https://twitter.com/

[5] https://www.facebook.com/

[6] https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

of Q1 2020 [7]. In fact, social networks have proven to be extremely useful for generating news, especially in crisis, due to their inherent ability to spread breaking news much more quickly than traditional media [7].

Fake news has received enormous attention from the academic community because it can be created and published online more quickly and cheaply than traditional media in several different platforms as newspapers and television. Also, several researchers suggest that humans tend to seek out, consume, and create information that is aligned with their ideological beliefs, often resulting in the perception and exchange of fake information in the same way as like-minded communities [20]. In this paper, we describe our submission as part of our participation at PAN at CLEF 2020, and as Pothast et al. [14] established, this paper closes a cycle by supplying the motivation for the tackled problem, high-level descriptions of the courses of action taken, and the interpretation of the results obtained. In particular, this year the Profiling Fake News Spreaders on Twitter task is presented, where the main objective is to identify possible spreaders of fake news on social networks as a first step to prevent the spread of fake news among online users. Our main contributions are related to the statistical analysis of the language use of the fake news spreader profiles, having the hypothesis that these profiles are created mainly to spread negative opinions in the social networks. To do this, we use the central tendency metrics (mean, median and mode), the use of polar and emotion classification and a vector of processed words thinking that these classifiers become a contributing factor in finding those features of the fake news spreader profile.

The rest of the paper is structured as follows. Section 2 introduces the related work. Section 3 describes the data set used in our strategy for celebrity characterization. Section 4 presents the details of the proposed strategy. Section 5 and 6 discuss the analysis of specific features and evaluation results. Finally, Section 7 presents some remarks and future work.

2   Related work

Profiling fake news broadcasters and detecting fake news are among the most complex tasks amidst natural language processing tasks. In addition, social media sites such as Facebook and Twitter are among the largest sources of news dissemination networks [2,5,22]. The detection of fake news is an activity that in recent years has generated great importance in different areas of society, as a phenomenon that is constantly growing. In this section we review some of the most recent work published.

Fake news detection has been studied from different approaches and techniques according to the scope and format of the available fake news data [11,17,9]. The most recent works are oriented towards using dynamic models of languages as those proposed by exBAKE [8] that mitigates the problem of data imbalance. Similarly, Cui et al [4] propose a deep end-to-end architecture which alleviates the heterogeneity introduced by multimodal data and it better captures the representation of user sentiment, as well. Rangel et al.[15] propose a Low Dimensionality Representation (LDR) model to reduce the possible over-fitting for identifying the language variation of different

---

[7] https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Spanish-speaking countries, which may help discriminate among different types of authors.

In general, current approaches based on deep neural networks have been successful in detecting Fake news. Still, there are other types of investigations that use traditional techniques sort of term frequency-inverse document frequency (TF-IDF), part-of-speech (POS) tagging, n-grams, among others. In relation to the TF-IDF approaches, we highlight the research of Ahmed et al.[1] who used a Stochastic Gradient Descent model using TF-IDF from the bi-grams. With regard to the part-of-speech (POS) tagging approach, the results presented by Rubin et al.[10] are highlighted. They used bi-grams with POS tagging to determine whether a news item was fake or not. Wynne et al.[21] propose a fake news detection system that considers the content of online news articles through the use of the word n-grams and the analysis of n-grams characters. Shu et al.[19] analyses the correlation between user profiles and fake news extracting implicit and explicit linguistic characteristics using a Linear Regression model, the use of metrics and The Five-Factor Model (FFM) unsupervised classification model for personality prediction. Finally, Giachanou et al.[6] improve the performance of their classification model CheckerOrSpreader for user profiles as potential fact checker or a potential fake news spreader combining a Convolutional Neural Network (CNN), The Five-Factor Model (FFM) prediction model with word embedding, and the LIWC software for tracking language patterns.

3 Materials and Methods

3.1 Data Description

The data set for task Profiling Fake News Spreaders on Twitter at PAN 2020 consists of 300 user profiles that spread fake news on social media. For this, files in XML format were provided with the content of 100 associated tweets of each author; this set includes texts in Spanish and English.

3.2 Model Description

In this section we describe the predictive model used in our submission. The model used for the task of Profiling Fake News Spreaders on Twitter. Figure 1 shows the description model.

3.3 Resources

To extract emotion and polarity from each comment associated with a user profile from the dataset, the NRC Emotion Lexicon [12] and a Combined Spanish Lexicon (CSL) [13] were used. The NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust), and two sentiments (negative and positive); it also includes translations in over 100 languages. The annotations were manually done by crowdsourcing. A Combined Spanish Lexicon is an approach for sentiment analysis includes an ensemble of six lexicons in Spanish and a weighted bag of words strategy.
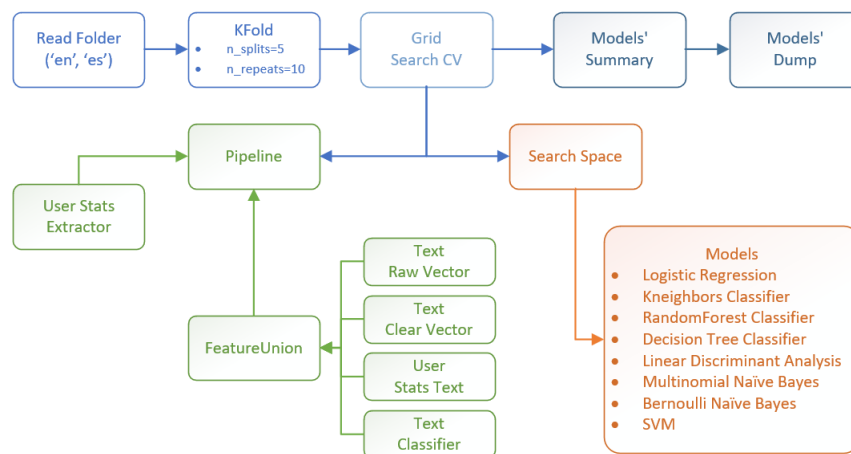
Figure 1. Model for the task of profiling fake news spreaders.

3.4 Preprocessing

Initially, a cleaning and pre-processing process is applied to the texts of the 300 users. In this way, the resulting corpus is ready for both languages and integrated into a feature vector.

Then, we applied a processing pipeline using Scikit-Learn to create new text features. Later, the GridSearchCV library was used to make a better search taking into account hyperparameters of various previously configured classifiers.

3.5 Feature Extraction

The first part of the pipeline was in charge of reading the text in both languages. Later, a feature vector of these texts was created and then a final preprocessing per individual was performed, resulting in a feature vector associated to each one of them, which sought to analyze the frequencies of the text's features such as emojis, emoticons, hashtags, URL or mentions.

A polarity analysis was carried out for each individual, taking into account the polarity of each of the messages shared by this user on the social network Twitter. Then, the amount of negative or positive comments was averaged, seeking to support the hypothesis that suggests a correct identification of fake profiles could occur through an analysis of the polarity on their messages since there is a correlation between a fake user and the negative polarity of the content shared on the network.

In the same scenario, the calculation of emotions for each of the texts was done by means of a lexicon of emotions, which allowed to identify if the emotions were binding characteristics of fake content.

3.6 Settings and Classifiers.

Emotional and polarity results, as well as statistics of the individual, were integrated into a single vector of characteristics to implement the classification model later. This model comprised a set of classifiers (Logistic Regression, K-Neighbors Classifier, Random Forest Classifier, Decision Tree Classifier, Linear Discriminant Analysis LDA, Multinomial Naive Bayes, Bernoulli Naive Bayes and Super Vector Machine) with which the hyperparameters were configured.

The hyperparameters tuning goal was to search a classifier with the best performance associated that could have been generated for each of the reports by GridSearchCV library and taking into account the pipeline.

The obtained results showed that the best performance was found using Random Forest with an accuracy of 76% for Spanish and 71.7% for English. This performance did not require changes on the settings for each language.

Finally, it is worth mentioning that the pipeline allowed to generate the classifiers, save them, serialize the pipeline with the classifiers and materialize them to perform the final execution of the model.

4 Experiments and Analysis of Results

As presented in Table 1, the summary shows the performance of the dummy profiles calculated for the challenge. For the class of dummy profiles, you can notice the best classification model, the accuracy obtained with it, and the characteristics that best worked for the classification. The classifier with the best performance was Random Forest. Furthermore, there is a union of the characteristics coming from raw text, cleaned text, the text statistics by profile and the polarity and the emotion classification of each tweet. Finally, a features vector is created with the objective of grouping the profile's language and sociolinguistic characteristics.

Table 1. Summary of results in the task of profiling fake news spreaders

Model	Accuracy (es)	Accuracy (en)
Logistic Regression	0.643	0.650
Kneighbors Classifier	0.640	0.577
RandomForest Classifier	**0.780**	**0.737**
Decision Tree Classifier	0.597	0.673
Linear Discriminant Analysis	NaN	NaN
Multinomial Naïve Bayes	NaN	NaN
Bernoulli Naïve Bayes	0.680	0.670
SVM	0.630	0.677

4.1 Baselines

Table 2 represents the predicted accuracy of our model for both languages compared to the baseline models made by the members in charge of the task. The main results show that the SYMANTO (LSDE) and SVM + c nGrams models outperform our model with an average difference of 4.5% and 1.3%, respectively. It should be noted that our performance is better in the English language concerning the SVM + c nGrams ; however, the performance drops if the analysis is in the Spanish language. On the contrary, our model has a better performance than the other models with a wide difference of 21.8% for the RANDOM baseline model and 2.8% with the closest baseline model.

Table 2. Performance of the different models on PAN at CLEF

Model	Accuracy(En)	Accuracy(ES)	Accuracy(Avg)
SYMANTO (LDSE)	0.745	0.790	0.768
SVM + c nGrams	0.680	0.790	0.735
morenosandoval20	**0.715**	**0.730**	**0.723**
NN + w nGrams	0.690	0.700	0.695
EIN	0.640	0.640	0.640
LSTM	0.560	0.600	0.580
RANDOM	0.510	0.500	0.505

On the other hand, if we compare our results with the performance of Ghanem et al. (2020) in the identification of fake news in Twitter, the general performance of the model carried out by the authors in the English language is far below 6.5%; however, the main class *clickbait* has a better performance than ours by a difference of 24.5%.

5 Discussion and Conclusion

The task of Profiling Fake News Spreaders on Twitter PAN at CLEF 2020 generated several challenges that are worth highlighting.

The collection and analysis of other language-related elements are of implicit context for this task of profiling fake news spreaders. Therefore, identifying profiles from their texts is an interesting approach where we can observe the analysis of variables in the use of some words that denote the social use of "sociolect" or "idiolect" languages. Therefore, this collection enables profile the own features of a specific language allow to increasing the accuracy in this type of natural language processing task.

This study associates text-based statistics with the length of characters and with the use of symbols, emojis, and expressions such as hashtags that can indicate semiotics. Texts are also used to make comments to other users by creating mentions within the network and finally referring to external sources of information in the URLs that can guide or give context to the messages. These messages imply different measurements

than the use of lexical or syntactic characteristics. By studying text-based statistics and other psychographic characteristics, such as emotion and polarity, it is possible to improve the precision of the classification processes on demographic, sociological, psychographic, and behavioral variables of fake news spreaders on Twitter.

Acknowledgements

References

1. Ahmed, H., Traore, I., Saad, S.: Detection of online fake news using n-gram analysis and machine learning techniques. In: International conference on intelligent, secure, and dependable systems in distributed and cloud environments. pp. 127–138. Springer (2017)
2. Ahmed, H., Traore, I., Saad, S.: Detecting opinion spams and fake news using text classification. Security and Privacy 1(1), e9 (2018)
3. Bondielli, A., Marcelloni, F.: A survey on fake news and rumour detection techniques. Information Sciences 497, 38–55 (2019)
4. Cui, L., Wang, S., Lee, D.: Same: sentiment-aware multi-modal embedding for detecting fake news. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. pp. 41–48 (2019)
5. Ghanem, B., Rosso, P., Rangel, F.: An emotional analysis of false information in social media and news articles. ACM Trans. Internet Technol. 20(2) (Apr 2020), https://doi.org/10.1145/3381750
6. Giachanou, A., Ríssola, E., Ghanem, B., Crestani, F., Rosso, P.: The Role of Personality and Linguistic Patterns in Discriminating Between Fake News Spreaders and Fact Checkers, pp. 181–192 (06 2020)
7. Imran, M., Castillo, C., Diaz, F., Vieweg, S.: Processing social media messages in mass emergency: Survey summary. In: Companion Proceedings of the The Web Conference 2018. pp. 507–511 (2018)
8. Jwa, H., Oh, D., Park, K., Kang, J.M., Lim, H.: exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (bert). Applied Sciences 9(19), 4062 (2019)
9. Kochkina, E., Liakata, M., Augenstein, I.: Turing at semeval-2017 task 8: Sequential approach to rumour stance classification with branch-lstm. arXiv preprint arXiv:1704.07221 (2017)
10. Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., et al.: The science of fake news. Science 359(6380), 1094–1096 (2018)
11. Long, Y.: Fake news detection through multi-perspective speaker profiles. Association for Computational Linguistics (2017)

12. Mohammad, S.M., Turney, P.D.: Crowdsourcing a word–emotion association lexicon. Computational Intelligence 29(3), 436–465 (2013)
13. Moreno-Sandoval, L.G., Beltrán-Herrera, P., Vargas-Cruz, J.A., Sánchez-Barriga, C., Pomares-Quimbaya, A., Alvarado-Valencia, J.A., García-Díaz, J.C.: Csl: A combined spanish lexicon - resource for polarity classification and sentiment analysis. In: Proceedings of the 19th International Conference on Enterprise Information Systems - Volume 1: ICEIS,. pp. 288–295. INSTICC, SciTePress (2017)
14. 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)
15. 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)
16. 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-WS.org (Sep 2020)
17. Ruchansky, N., Seo, S., Liu, Y.: Csi: A hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 797–806 (2017)
18. Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., Liu, Y.: Combating fake news: A survey on identification and mitigation techniques. ACM Transactions on Intelligent Systems and Technology (TIST) 10(3), 1–42 (2019)
19. 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 Computer Society, Los Alamitos, CA, USA (apr 2018), https://doi.ieeecomputersociety.org/10.1109/MIPR.2018.00092
20. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter 19(1), 22–36 (2017)
21. Wynne, H.E., Wint, Z.Z.: Content based fake news detection using n-gram models. In: Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services. pp. 669–673 (2019)
22. Zhou, X., Zafarani, R.: Fake news detection: An interdisciplinary research. In: Companion Proceedings of The 2019 World Wide Web Conference. pp. 1292–1292 (2019)