

Document downloaded from:

<http://hdl.handle.net/10251/178922>

This paper must be cited as:

Ghanem, BHH.; Ponzetto, SP.; Rosso, P. (2020). FacTweet: Profiling Fake News Twitter Accounts. Springer. 35-45. https://doi.org/10.1007/978-3-030-59430-5_3



The final publication is available at

https://doi.org/10.1007/978-3-030-59430-5_3

Copyright Springer

Additional Information

FacTweet: Profiling Fake News Twitter Accounts

Bilal Ghanem¹, Simone Paolo Ponzetto², and Paolo Rosso¹

¹ Universitat Politècnica de València, Spain
{bigha@doctor., proso@dsic.}jupv.es

² University of Mannheim, Germany
simone@informatik.uni-mannheim.de

Abstract. We present an approach to detect fake news in Twitter at the account level using a neural recurrent model and a variety of different semantic and stylistic features. Our method extracts a set of features from the timelines of news Twitter accounts by reading their posts as chunks, rather than dealing with each tweet independently. We show the experimental benefits of modeling latent stylistic signatures of mixed fake and real news with a sequential model over a wide range of strong baselines.

Keywords: Fake News · Twitter Accounts · Factual Accounts

1 Introduction

Social media platforms have made the spreading of fake news easier, faster as well as able to reach a wider audience. Social media offer another feature which is the anonymity for the authors, and this opens the door to many suspicious individuals or organizations to utilize these platforms. Recently, there has been an increased number of spreading fake news and rumors over the web and social media [23]. Fake news in social media vary considering the intention to mislead. Some of these news are spread with the intention to be ironic or to deliver the news in an ironic way (satirical news). Others, such as propaganda, hoaxes, and clickbaits, are spread to mislead the audience or to manipulate their opinions. In the case of Twitter, suspicious news annotations should be done on a tweet rather than an account level, since some accounts mix fake with real news. However, these annotations are extremely costly and time consuming – i.e., due to high volume of available tweets. Consequently, a first step in this direction, e.g., as a pre-filtering step, is the task of detecting fake news at the account level. The main obstacle for detecting suspicious Twitter accounts is due to the behavior of mixing some real news with the misleading ones. Consequently, we investigate a way to detect suspicious accounts by considering their tweets in groups (chunks). Our hypothesis is that suspicious accounts have a unique pattern in posting tweet sequences. Since their intention is to mislead, the way they transition from one set of tweets to the next has a hidden signature, biased by their intentions. Therefore, reading these tweets in chunks has the potential to improve the detection of the fake news accounts.

In this work, we investigate the problem of discriminating between factual and non-factual accounts in Twitter. To this end, we collect a dataset of tweets using a list of *propaganda*, *hoax* and *clickbait* accounts and compare different versions of sequential chunk-based approaches using a variety of feature sets against several baselines. Several approaches have been proposed for news verification, whether in social media (rumors detection) [23,22], or in news claims [4]. The main line of research of previous works is to verify the textual tweets but not their sources. Another existing direction in the literature is the detection of online trolls or bots [21]. This is different from our setting, since online trolls are less formal and try to imitate individuals by spreading a mixed content, e.g., social media funneling [6], news, personal opinions [8], etc.. On the other hand, the content of fake news Twitter accounts is formal, objective, and focused on spreading news content only. To the best of our knowledge, this is the first work aiming to detect factuality at account level, specifically from a textual perspective. The contributions of this work are the following ones:

- We propose an approach to detect non-factual Twitter accounts by treating post streams as a sequence of tweets’ chunks. We test several semantic and dictionary-based features together with a neural sequential approach, and apply an ablation test to investigate their contribution.
- We benchmark our approach against other approaches that discard the chronological order of the tweets or read the tweets individually. The results show that our approach produces superior results at detecting non-factual accounts.

The rest of the paper is structured as follows. In the following section, we present an overview on the related work. In Section 3, we present the methodology of our approach. Section 4 describes the collected dataset, the experiments, and the results. Finally, we draw some conclusions and discuss possible future works.

2 Related Work

Fake news detection has gained a lot of attention and has been approached from several perspectives in both social media and online news sites. Our work is closely related to the following areas.

2.1 Fake News Sources

Previous works focuses on approaching and analyzing online news texts or claims [15,16]. Instead, the work in [4] looks at characterizing entire news media. The authors propose a set of features for the detection of low-factual news media. They use features based on Wikipedia pages and Twitter accounts, like *Does it have Wikipedia page?*, *Is the Twitter account verified?*, etc.. Also, they use manual features to identify the low-factual media using their malicious URLs,

a set of features to capture the reporting language of the news articles, and the *Alexa Rank* metric to model the web traffic over the news media. The system shows a macro-F1 value of ~ 0.6 over 3 classes, low, mixed, and high factuality. Another work [5] approaches the problem of detecting the trustworthiness of news media by combining the factuality with bias in a multi-task ordinal regression framework that models the two problems jointly. The authors use the same feature set that was proposed in [4] and show that their system can generate a good result using the Mean Absolute Error metric with a value of ~ 0.53 . In the direction of understanding the characteristics of not credible news sources, the work [1] studied the correlation of a set of features with credible and transparent news media. And in [2], the same authors propose a regression task for source credibility assessment using a set of features like *Google page rank*, *Alexa rank*, *Spam score*, etc., and achieve a value of ~ 17.7 using RMSE (Root Mean Squared Error).

2.2 Fishy Twitter Accounts

Suspicious accounts in social media play a key role in spreading fake news and deceiving other online users. A set of works has been done to detect bots or trolls accounts. Many works [6,3,18,12] propose a set of features to detect online trolls, starting from textual features such as *the existence of hashtags and URLs in the trolls tweets*, *bag-of-words*, *part-of-speech features* or with including more sophisticated features such as *bot likelihood*, *topic-based information*, and *activity-related account metadata*. The majority of these works focus on online Russian trolls that were spreading fake news during the US 2016 elections, and produced superior results comparing to baselines.

The work in [9] propose a bots detection system called *BotorNot*³ to detect bots in Twitter. The system uses content, sentiment, friend, network, temporal, and user features. The authors use a dataset of Twitter accounts – collected previously in another work – that spread tweets about online products (advertisements), duplicate others’ tweets, etc.. The system obtained an Area Under ROC Curve (AUC) value of 0.95. In a similar attempt, the authors of [11] propose *SentiBot* to detect online bots in the context of the 2014 Indian election. The system uses a large combination of features that contain sentiment, topic, network, and syntax features. The proposed model obtains Receiver Operating Characteristic Curve (ROC) value of ~ 0.73 on a dataset collected within a year from Twitter.

3 Methodology

Given a news Twitter account, we read its tweets from the account’s timeline. Then we sort the tweets by the posting date in ascending way and we split them

³ Later on, the authors created an online API for the system called Botometer in: <https://botometer.iuni.iu.edu>.

into N chunks. Each chunk consists of a sorted sequence of tweets labeled by the label of its corresponding account. We extract a set of features from each chunk and we feed them into a recurrent neural network to model the sequential flow of the chunks’ tweets. We use an attention layer with dropout to attend over the most important tweets in each chunk. Finally, the representation is fed into a softmax layer to produce a probability distribution over the account types and thus predict the factuality of the accounts. Since we have many chunks for each account, the label for an account is obtained by taking the majority class of the account’s chunks.

Input Representation. Let t be a Twitter account that contains m tweets. These tweets are sorted by date and split into a sequence of chunks $ck = \langle ck_1, \dots, ck_n \rangle$, where each ck_i contains s tweets. Each tweet in ck_i is represented by a vector $v \in \mathbb{R}^d$, where v is the concatenation of a set of features’ vectors, that is $v = \langle f_1, \dots, f_n \rangle$. Each feature vector f_i is built by counting the presence of tweet’s words in a set of lexical lists.

Features. We argue that different kinds of features like the sentiment of the text, morality, and other text-based features are critical to detect the nonfactual Twitter accounts by utilizing their occurrence during reporting the news in an account’s timeline. We employ a rich set of features borrowed from previous works in fake news, bias, and rumors detection [23,22,4].

- **Emotion:** We build an emotions vector using word occurrences of 8 emotion types from the NRC lexicon [20], which contains ~ 14 K words labeled using the eight Plutchik’s emotions. The emotions feature can detect if an account is frequently triggering negative emotions like fear, anger, etc.
- **Sentiment:** We extract the sentiment of the tweets by employing Effect-WordNet [7], SenticNet⁴, NRC [20]⁵, and subj_lexicon [24], where each has the two sentiment classes, *positive* and *negative*. The sentiment feature can highlight the polarity in a more abstract level than emotions.
- **Morality:** Features based on morality foundation theory [17] where words are labeled in one of the following 10 categories (*care, harm, fairness, cheating, loyalty, betrayal, authority, subversion, sanctity, and degradation*). Using the morality features, we can highlight if some Twitter fake news accounts are posting more frequently news about harmful, subversion, or degradation events. It has been proved that fake news accounts usually post messages about very negative events to catch the readers’ eyes [15].
- **Style:** We use canonical stylistic features, such as the count of question marks, exclamation marks, consecutive characters and letters⁶, links, hash-tags, users’ mentions. In addition, we extract the uppercase ratio and the tweet length. We aim to detect if a specific account uses a fixed language style.

⁴ <https://sentic.net/>

⁵ NRC has also two sentiment categories, positive and negative.

⁶ We considered 2 or more consecutive characters, and 3 or more consecutive letters.

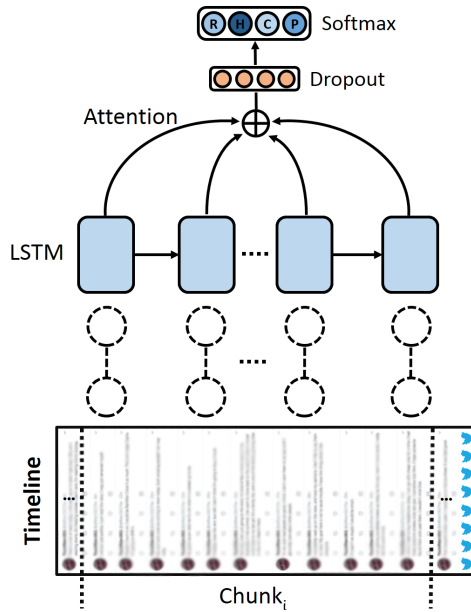


Fig. 1. The FacTweet’s architecture.

- **Words embeddings:** We extract words embeddings of the tweets’ words using *Glove840B – 300d*⁷ pretrained model⁸. The tweet final representation is obtained by averaging its words embeddings. The word embeddings is important to extract the topic information from the messages. Fake news accounts usually post news about specific topics. Also, this feature is complementary to the previous ones where, for an example, detecting a negative sentiment without knowing the topic of the messages would not be useful.

Model. To account for chunk sequences we make use of a *de facto* standard approach and opt for a recurrent neural model using long short-term memory (LSTM). In our model, the sequence consists of a sequence of tweets belonging to one chunk (Figure 1). The LSTM learns the hidden state h_t by capturing the sequential changes in the timesteps. The produced hidden state h_t at each time step is passed to the attention layer which computes a ‘context’ vector c_t as the weighted mean of the state sequence h by: $c_t = \sum_{j=1}^T \alpha_{tj} h_j$, Where T is the total number of timesteps in the input sequence and α_{tj} is a weight computed at each time step j for each state h_j .

⁷ <https://nlp.stanford.edu/projects/glove/>

⁸ Experimentally, we found that the *GloVe* model achieves better results than *Google News word2vec* or *fastText* models.

Table 1. Statistics on the data with respect to each account type: propaganda (**P**), clickbait (**C**), hoax (**H**), and real news (**R**).

	Accounts Types			
	P	C	H	R
# of accounts	96	36	7	32
Max # of tweets/account	3,250	3,246	3,250	3,250
Min # of tweets/account	33	877	453	212
Avg # of tweets/account	2,978	3,112	2,723	3,124
Total # of tweets	291,885	112,050	19,065	99,967

4 Experiments and Results

Data. We build a dataset of Twitter accounts based on two lists annotated by professional journalists. For the non-factual accounts, we rely on a list of approximately 180 Twitter accounts from [22]⁹. This list was created based on public resources¹⁰ where suspicious Twitter accounts were annotated with the main fake news types (clickbait, propaganda, satire, and hoax). We discard the satire labeled accounts since their intention is not to mislead or deceive. On the other hand, for the factual accounts, we use a list with another 32 Twitter accounts from [19] that are considered trustworthy by independent third parties¹¹. We discard accounts that publish news in languages other than English (e.g., Russian or Arabic). Moreover, to ensure the quality of the data, we remove the duplicate, media-based, and link-only tweets. For each account, we collect the maximum amount of tweets allowed by Twitter API. Table 1 presents statistics on our dataset.

Baselines. We compare our approach (FacTweet) to the following baselines:

- **LR + Bag-of-words:** We aggregate the tweets of a feed and we use a bag-of-words representation with a logistic regression (LR) classifier.
- **Tweet2vec:** We use the model proposed in [10] which is a Bidirectional Gated recurrent neural network to predict the tweets based on their hashtags. Their model converts the tweets into character one-hot encoding and feed them to the model. We used our collected dataset which consists of ~ 0.5 M tweets to train this model. We keep the default parameters that were provided with the implementation. To represent the tweets, we use the decoded embedding produced by the model. With this baseline we aim at assessing if the tweets’ hashtags may help detecting the non-factual accounts.
- **LR + All Features (tweet-level):** We extract all our features from each tweet and feed them into a LR classifier. Here, we do not aggregate over tweets and thus view each tweet independently.

⁹ Many of the accounts were deactivated during the collecting process, consequently only 144 accounts were used.

¹⁰ <http://www.propornot.com/p/the-list.html>

¹¹ <https://tinyurl.com/yctvve9h>

- **LR + All Features (chunk-level)**: We concatenate the features’ vectors of the tweets in a chunk and feed them into a LR classifier.
- **FacTweet (tweet-level)**: Similar to the FacTweet approach, but at tweet-level; the sequential flow of the tweets is not utilized. We aim at investigating the importance of the sequential flow of tweets.
- **Botometer**: We use Botometer [9], a state-of-the-art Twitter bots detection system. Botometer uses Network, User, Friends, Temporal, Content, and Sentiment features for bots detection. We aim at checking whether we can detect the Twitter fake news accounts using a bots detection system, where such accounts might have employed automated softwares to release fake news. Also, with this baseline, we assess the performance of the state-of-the-art bots detection system in our task. We fed the Botometer generated predictions to a Random Forest (RF) classifier. We chose RF after testing several classifiers, e.g., Logistic Regression, Support Vector Machine, Naive Bayes, and Feed Forward Neural Network.
- **Top- k replies, likes, or re-tweets**: Some approaches in rumors detection use the number of replies, likes, and re-tweets to detect rumors [13]. Thus, we extract top k replied, liked or re-tweeted tweets from each account to assess the accounts factuality. We tested different k values between 10 tweets to the max number of tweets from each account. Figure 2 shows the macro-F1 values for different k values. It seems that $k = 500$ for the top *replied* tweets achieves the highest result. Therefore, we consider this as a baseline.

Experimental Setup. We report the results using accuracy and macro F1. We experiment with 25% of the accounts for validation and parameters selection, and we apply 5 cross-validation on the rest of the data (75%). The validation split is extracted on the class level using stratified sampling: for this, we take a random 25% of the accounts from each class since the dataset is unbalanced. Discarding the classes’ size in the splitting process may affect the minority classes (e.g., hoax). We use hyperopt library¹² to select the hyper-parameters on the following values: LSTM layer size (16, 32, 64), dropout (0.0 – 0.9), activation function (*relu*, *selu*, *tanh*), optimizer (*sgd*, *adam*, *rmsprop*) with varying the value of the learning rate (1e-1,...,1e-5), and batch size (4, 8, 16). To reduce the effect of overfitting in FacTweet, we use the early stopping technique. For the baselines’ classifier, we tested many classifiers and the LR showed the best overall performance.

Results. Table 2 presents the results. We present the results using a chunk size of 20, which was found to be the best size using the validation set. Figure 3 shows the results of different chunks sizes.

FacTweet performs better than the proposed baselines and obtains the highest macro-F1 value of 0.565. Our results indicate the importance of taking into account the sequence of the tweets in the accounts’ timelines. The sequence of these tweets is better captured by our proposed model sequence-agnostic or non-neural classifiers. Moreover, the results demonstrate that the features at

¹² <https://github.com/hyperopt>

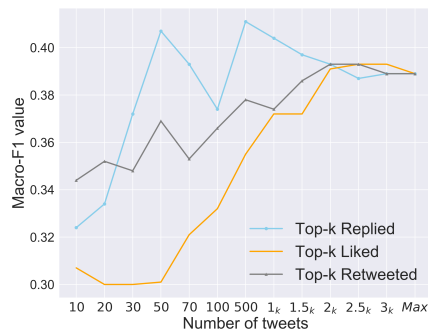


Fig. 2. Results on the top-K replied, linked or re-tweeted tweets.

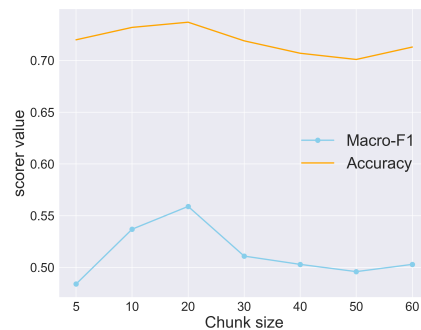


Fig. 3. The FacTweet performance on difference chunk sizes.

Table 2. Results on accounts classification.

Methods	A	P	R	F1
Baselines				
Majority Class	0.563	0.141	0.251	0.18
Random class	0.252	0.21	0.21	0.209
Bag-of-Words	0.601	0.252	0.327	0.284
Tweet2vec	0.558	0.157	0.213	0.181
Botometer	0.512	0.356	0.371	0.363
Tweet-level approaches				
LR + All	0.671	0.378	0.411	0.393
LR + All (top-500 replied)	0.443	0.368	0.467	0.411
LR + FacTweet	0.651	0.34	0.37	0.351
Chunk-level approaches				
LR + All	0.737	0.603	0.552	0.559
FacTweet	0.74	0.549	0.582	0.565

tweet-level do not perform well to detect the Twitter accounts factuality, since they obtain a result near to the majority class (0.18). Another finding from our experiments shows that the performance of the Tweet2vec is weak. This demonstrates that tweets' hashtags are not informative to detect non-factual accounts. Furthermore, the results show that the performance of the Botometer system is weak comparing to the other models, and this emphasizes that fake news accounts use more advanced techniques to spread fake news comparing to the more basic bots techniques. Also, we argue that the low performance of Botometer is due to the different nature of our task. Bots and trolls spread mixed information that contains advertisements and opinions, where the proposed bots detection systems, like Botometer, utilize features that give importance to such information in tweets. Also, bots accounts usually are not well connected with other users accounts (considering the network features e.g. number of followers), and such features are important to detect these accounts but not fake news accounts

that gained the trust of many followers. In Table 3, we present ablation tests so as to quantify the contribution of subset of features. The results indicate that most performance gains come from words embeddings, style, and morality features. Other features (emotion and sentiment) show lower importance: nevertheless, they still improve the overall system performance (on average 0.35% macro-F₁ improvement). These performance figures suggest that non-factual accounts use semantic and stylistic hidden signatures mostly while tweeting news, so as to be able to mislead the readers and behave as reputable (i.e., factual) sources.

Table 3. Ablation tests.

Methods	Accuracy	Precision	Recall	F1
LR + All	0.737	0.603	0.552	0.559
– Emotion	0.731	0.581	0.535	0.557
– Sentiment	0.731	0.535	0.575	0.554
– Morality	0.725	0.554	0.542	0.548
– Style	0.737	0.521	0.508	0.514
– Words embeddings	0.678	0.43	0.444	0.437

Since the dataset is highly imbalanced, we apply upsampling by replicating the minority classes. In Table 4 we present the results. For the model (LR + All) that is applied on the chunk-level, we do not get any improvement. For the FacTweet, we notice a small improvement in terms of F1 score. We leave a more fine-grained, diachronic analysis of semantic and stylistic features – how semantic and stylistic signature evolve across time and change across the accounts’ timelines – for future work.

Table 4. Up-sampling (Up-s).

Methods	Accuracy	F1 _{macro}
LR + All	0.737	0.559
LR + All + Up-s	0.737	0.559
FacTweet	0.74	0.565
FacTweet + Up-s	0.74	0.571

5 Conclusions and Future Work

In this paper, we proposed a model that utilizes chunked timelines of tweets and a recurrent neural model in order to infer the factuality of a Twitter news account. Our experimental results indicate the importance of analyzing tweet stream into chunks, as well as the benefits of heterogeneous knowledge source (i.e., lexica as well as text) in order to capture factuality. In future work, we would like to extend

this line of research with further in-depth analysis to understand the flow change of the used features in the accounts' streams. Moreover, we would like to take our approach one step further incorporating explicit temporal information, e.g., using timestamps. Crucially, we are also interested in developing a multilingual version of our approach, for instance by leveraging the now ubiquitous cross-lingual embeddings [14]. Finally, we will investigate the potential of applying transfer learning from social media posts. As transfer learning models are starving for data, we will work on extending the used dataset with further social media accounts to enable more accurate fine-tuning process.

Acknowledgment

The work of Paolo Rosso was partially funded by the the Spanish MICINN under the research project MISMIS-FAKEHATE on Misinformation and Miscommunication in social media: FAKE news and HATE speech (PGC2018-096212-B-C31).

References

1. Aker, A., Kevin, V., Bontcheva, K.: Credibility and Transparency of News Sources: Data Collection and Feature Analysis. In: Arxiv (2019)
2. Aker, A., Kevin, V., Bontcheva, K.: Predicting News Source Credibility. In: Arxiv (2019)
3. Badawy, A., Lerman, K., Ferrara, E.: Who Falls for Online Political Manipulation? In: Companion Proceedings of The 2019 World Wide Web Conference. pp. 162–168. ACM (2019)
4. Baly, R., Karadzhov, G., Alexandrov, D., Glass, J., Nakov, P.: Predicting Factuality of Reporting and Bias of News Media Sources. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 3528–3539 (2018)
5. Baly, R., Karadzhov, G., Saleh, A., Glass, J., Nakov, P.: Multi-Task Ordinal Regression for Jointly Predicting the Trustworthiness and the Leading Political Ideology of News Media. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 2109–2116 (2019)
6. Boyd, R.L., Spangher, A., Fourney, A., Nushi, B., Ranade, G., Pennebaker, J., Horvitz, E.: Characterizing the Internet Research Agency's Social Media Operations During the 2016 US Presidential Election using Linguistic Analyses. In: PsyArXiv (2018)
7. Choi, Y., Wiebe, J.: +/-EffectWordNet: Sense-level Lexicon Acquisition for Opinion Inference. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1181–1191 (2014)
8. Clark, E.M., Williams, J.R., Jones, C.A., Galbraith, R.A., Danforth, C.M., Dodds, P.S.: Sifting Robotic from Organic Text: a Natural Language Approach for Detecting Automation on Twitter. *Journal of Computational Science* **16**, 1–7 (2016)

9. Davis, C.A., Varol, O., Ferrara, E., Flammini, A., Menczer, F.: Botornot: A System to Evaluate Social Bots. In: Proceedings of the 25th International Conference Companion on World Wide Web. pp. 273–274. International World Wide Web Conferences Steering Committee (2016)
10. Dhingra, B., Zhou, Z., Fitzpatrick, D., Muehl, M., Cohen, W.W.: Tweet2Vec: Character-Based Distributed Representations for Social Media. In: The 54th Annual Meeting of the Association for Computational Linguistics (ACL). p. 269 (2016)
11. Dickerson, J.P., Kagan, V., Subrahmanian, V.: Using Sentiment to Detect Bots on Twitter: Are Humans more Opinionated than Bots? In: 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014). pp. 620–627. IEEE (2014)
12. Ghanem, B., Buscaldi, D., Rosso, P.: TexTrolls: Identifying Russian Trolls on Twitter from a Textual Perspective. arXiv preprint arXiv:1910.01340 (2019)
13. Ghanem, B., Cignarella, A.T., Bosco, C., Rosso, P., Rangel, F.: UPV-28-UNITO at SemEval-2019 Task 7: Exploiting Post’s Nesting and Syntax Information for Rumor Stance Classification. In: Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval). pp. 1125–1131 (2019)
14. Ghanem, B., Glavas, G., Giachanou, A., Ponzetto, S.P., Rosso, P., Pardo, F.M.R.: UPV-UMA at checkthat! lab: Verifying arabic claims using a cross lingual approach. In: Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12, 2019. (2019)
15. Ghanem, B., Rosso, P., Rangel, F.: An Emotional Analysis of False Information in Social Media and News Articles. *ACM Transactions on Internet Technology (TOIT)*. (2020)
16. 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. pp. 877–880 (2019)
17. Graham, J., Haidt, J., Nosek, B.A.: Liberals and Conservatives Rely on Different Sets of Moral Foundations. *Journal of Personality and Social Psychology* **96**(5), 1029 (2009)
18. Im, J., Chandrasekharan, E., Sargent, J., Lighthammer, P., Denby, T., Bhargava, A., Hemphill, L., Jurgens, D., Gilbert, E.: Still out there: Modeling and Identifying Russian Troll Accounts on Twitter. arXiv preprint arXiv:1901.11162 (2019)
19. Karduni, A., Wesslen, R., Santhanam, S., Cho, I., Volkova, S., Arendt, D., Shaikh, S., Dou, W.: Can You Verifi This? Studying Uncertainty and Decision-Making about Misinformation Using Visual Analytics. In: In Twelfth International AAAI Conference on Web and Social Media (ICWSM) (2018)
20. Mohammad, S.M., Turney, P.D.: Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. In: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. pp. 26–34 (2010)
21. Shao, C., Ciampaglia, G.L., Varol, O., Flammini, A., Menczer, F.: The Spread of Fake News by Social Bots. arXiv preprint arXiv:1707.07592 pp. 96–104 (2017)
22. Volkova, S., Shaffer, K., Jang, J.Y., Hodas, N.: Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL) (Volume 2: Short Papers). vol. 2, pp. 647–653 (2017)
23. Vosoughi, S., Roy, D., Aral, S.: The Spread of True and False News Online. *Science* **359**(6380), 1146–1151 (2018)

24. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (EMNLP) (2005)