Data & Knowledge Engineering 138 (2022) 101960

Contents lists available at ScienceDirect

Data & Knowledge Engineering

journal homepage: www.elsevier.com/locate/datak

The impact of psycholinguistic patterns in discriminating between fake news spreaders and fact checkers

Anastasia Giachanou [a,*], Bilal Ghanem [b], Esteban A. Ríssola [c], Paolo Rosso [d], Fabio Crestani [c], Daniel Oberski [a]

[a] Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands
[b] Department of Computer Science, University of Alberta, Edmonton, Canada
[c] Faculty of Informatics, Università della Svizzera italiana, Lugano, Switzerland
[d] PRHLT Research Center, Universitat Politècnica de València, Valencia, Spain

ARTICLE INFO

Keywords:
Fake news
Linguistic analysis
Misinformation

ABSTRACT

Fake news is a threat to society. A huge amount of fake news is posted every day on social networks which is read, believed and sometimes shared by a number of users. On the other hand, with the aim to raise awareness, some users share posts that debunk fake news by using information from fact-checking websites. In this paper, we are interested in exploring the role of various psycholinguistic characteristics in differentiating between users that tend to share fake news and users that tend to debunk them. Psycholinguistic characteristics represent the different linguistic information that can be used to profile users and can be extracted or inferred from users' posts. We present the CheckerOrSpreader model that uses a Convolution Neural Network (CNN) to differentiate between spreaders and checkers of fake news. The experimental results showed that CheckerOrSpreader is effective in classifying a user as a potential spreader or checker. Our analysis showed that checkers tend to use more positive language and a higher number of terms that show causality compared to spreaders who tend to use a higher amount of informal language, including slang and swear words.

1. Introduction

Fake news is not a recent phenomenon. However, social networks have given users the opportunity to publish and share content that can be propagated very quickly. The recent unprecedented dissemination of online disinformation is a threat for democracy and journalism. The negative consequences of online disinformation can be highlighted during a wide range of important events including political elections [1–3], referendums [4] and health related incidents [5]. For example, research has shown how medical misinformation can result to false treatment advice as happened in the case of Ebola [6]. In the political domain, several researchers have underlined the influence of fake news on elections and referendums including U.S. presidential elections and Brexit [2,4,7]. The most recent example is regarding the COVID-19 pandemic for which a huge amount of fake news and conspiracy theories were propagated regarding its origin, transmission and treatment. In fact, the misleading information that was propagated prompted the World Health Organization to warn of an ongoing infodemic [8].

Regardless of all the attempts to raise awareness and help users learn to recognize fake content, there are still many users that believe them and share them online. In general, humans are not effective in recognizing if a piece of news is accurate or not. A study has showed that humans' ability to detect fake news is slightly better than chance [9]. This is due to a number of different reasons. First, fake news is written in a way to confuse users and to make them believe them by mixing true and false information. Another

* Corresponding author.
E-mail address: a.giachanou@uu.nl (A. Giachanou).

https://doi.org/10.1016/j.datak.2021.101960
Received 17 March 2021; Accepted 22 November 2021
Available online 15 December 2021

reason is that users tend to believe and interpret information that confirms their prior beliefs or values, known as *confirmation bias* [10]. Some users also tend to believe perceptions that their peers share and believe, known as the *bandwagon effect* [11]. Those phenomena make the problem of fake news detection a big challenge.

In an attempt to raise awareness, several fact-checking sites have been developed. The fact-checking sites publish information regarding the credibility of online articles and provide explanations of their evaluations. Some of those websites (e.g., politifact,[1]) check whether political claims are valid or not, whereas others (e.g., snopes[2] leadstories,[3]) focus on checking the validity of news articles. The decision of whether an article or a claim is fake or not is taken by media experts and journalists that analyze the content of the article/claim. As a result of the emergence of the fact-checking site, a new group of users has emerged in social media. Those users, known as *fact-checkers* or *checkers* share posts about articles' factuality based on what has been published on the fact-checking websites.

With the advancements of generative models that are extensively used to produce fake content, users' role in halting fake news propagation is becoming critical. Although bots are also used to propagate fake news [12], the contribution of users in their propagation is also very important since they decide to intentionally or unintentionally share fake content [13] thus reinforcing it. Profiling users that are involved in the creation, propagation or halting of fake news in any way is important and can be further exploited by systems that aim to detect fake content. Several studies from social psychology and communication have tried to profile users that are in any way involved in the creation or propagation of fake news. For example, a study from Bronstein et al. [14] suggested that dogmatic individuals and religious fundamentalists were more likely to believe false news, whereas Stanley et al. [15] suggested that individuals less likely to engage effortful, deliberative, and reflective cognitive processes were more likely to believe that COVID-19 pandemic was a hoax.

A lot of attempts have been made to build models that can profile users that are involved in any stage of the fake news life cycle. Shu et al. [16] analyzed different features, such as registration time, and found that users that share fake news have more recent accounts than users who share real news, whereas Giachanou et al. [17] analyzed the profile and linguistic differences between conspiracy propagators and non-propagators and found that conspiracy propagators use a larger number of swear words. Different from those studies, we focus on profiling users that share fake news and those that debunk them and explore if it is possible to differentiate between them.

In this study we assume that checkers are likely to express themselves using a set of different linguistic characteristics compared to spreaders. We are interested in exploring if there are any psycholinguistic differences between users that tend to spread fake news (*spreaders*) and those that tend to debunk them (*checkers*). We explore a range of different characteristics including sentiment, emotions, linguistic patterns, personality traits, communication style and readability. Those characteristics are either calculated directly from the users' posts or inferred by the posts. In addition, we present CheckerOrSpreader, a model that aims to classify a user as a spreader or checker. CheckerOrSpreader is based on a Convolution Neural Network (CNN) and combines information based on the text and the psycholinguistic characteristics of the users' posts.

More specifically, we focus on the following research questions:

RQ1. Can psycholinguistic features help differentiate between fake news spreaders and fact checkers?
RQ2. Which psycholinguistic variables show a statistically significant univariate difference between fake news spreaders and fact checkers?
RQ3. What are the top-N psycholinguistic characteristics that discriminate fake news spreaders and fact checkers in a multivariable prediction?

The contributions of this article can be summarized as follows:

1. We present CheckerOrSpreader, a model that uses a wide range of psycholinguistic features to discriminate between fake news spreaders and fact checkers.
2. We present a dataset that contains tweets posted by spreaders and checkers that is built with the help of a fact-checking website[3].
3. We analyze the differences of the different psycholinguistic characteristics between the fake news spreaders and fact checkers.
4. We analyze the importance of the different psycholinguistic characteristics on the task of distinguishing fake news spreaders from checkers, among Twitter users who shared fake news.

This paper extends and complements our previous work on discriminating between spreaders and checkers (Giachanou et al. [18]) by providing a more complete presentation of the topic and a more thorough analysis of the results. It contains considerable extensions of the original paper:

1. We explore a wider range of psycholinguistic characteristics such as communication style, readability and an expanded list of linguistic patterns.
2. We study the statistical differences of the psycholinguistic characteristics between spreaders and checkers.
3. We study the importance of the different psycholinguistic characteristics on differentiating between spreaders and checkers.

[1] See https://www.politifact.com.
[2] See https://www.snopes.com.
[3] See https://leadstories.com/.

The rest of the paper is organized as follows. Section 2 discusses related work on fake news detection and the role of users in the problem. Next we present the process we followed to create the collection and the CheckerOrSpreader model in Section 3. Section 4 presents the experimental setup and Section 5 our results and analysis. Finally, Section 6 discusses the limitations and the ethical concerns regarding our study followed by the conclusions and future work in Section 7.

2. Related work

In this section we present related work on fake news detection and on the role of users in disinformation.

2.1. Fake news detection

Automatic detection of fake news is very challenging because there are many different and sophisticated ways to manipulate news. A number of studies have explored the impact of different patterns that can be extracted from the content of the article. To this end, researchers have analyzed and explored different patterns that range from simple term frequencies [19] to image level features [20] and from linguistic characteristics [21] to contextualized word embeddings [22]. A thorough survey on the topic of automatic fake news detection can be found in [23,24].

Early attempts in automatic fake news detection explored the impact of features that were extracted from the text of the article. To this end, Volkova et al. [25] analyzed fake and real tweets from a linguistic perspective and found that real news tweets contained significantly fewer subjective terms and harmful words compared to fake ones. Also, their analysis showed that satirical news contained more loyalty and less betrayal morals compared to propaganda. They also built a model that was trained on graph-based, cues words, and syntax level information that managed to perform better compared to the baselines. Zhou et al. [19] investigated a wide range of lexicon-level, syntax-level, semantic-level and discourse-level information extracted from news articles. Zhou et al. used those features to train a supervised machine learning framework that outperformed the baselines.

Sentiment and emotions that are expressed in posts or news have been extensively explored regarding their effectiveness in fake news detection. One reason that sentiment and emotions attracted such a huge amount of attention in the field of fake news detection is that previous research showed that deceptive claims tended to be expressed differently in terms of emotions from truth [26]. Vosoughi et al. [13] investigated differences between true and false news stories distributed on Twitter from 2006 to 2017 and showed that false stories tended to inspire fear, disgust, and surprise in replies whereas true stories expressed anticipation, sadness, joy, and trust. Giachanou et al. [27] proposed a model based on Long Short Term Memory (LSTM) and on emotional signals extracted from the text of the claims and showed that their model was useful for determining whether a claim is credible or not. Ghanem et al. [28] also proposed a system based on LSTM that leveraged emotions with the aim of classifying the type of the article as satire, hoax, propaganda, clickbaits or real news. Their study showed that different types of disinformation have different emotional patterns. Recently, Guo et al. [29] proposed a dual emotion-based fake news detection model to learn content and comment emotion features and their relationship for publishers and users, respectively.

Although textual information is very important, additional information from different modalities can also be useful and complementary for fake news detection. To this end, some researchers have tried to incorporate features extracted from images. The majority of the multimodal approaches combined information from text represented with word embeddings and information from images extracted with pre-trained models such as VGG-16 and VGG-19 [20,30–32]. Some of the studies incorporated also similarity information between the headline and the image [20,33]. Zlatkova et al. [34] explored a variety of features extracted from the claim (e.g., claim text), the image (e.g., URL domains) and the relationship between the two (e.g., embedding similarity) to detect fake claims about images. In their study, they found that URL domains play a very important role in predicting the factuality of a claim with respect to an image.

External knowledge has also proven to be very useful in the field of automatic fake news detection. In order to obtain external knowledge, researchers have used search engines to retrieve relevant documents from which they then extracted additional features [35,36]. To address the task of fake news detection, Popat et al. [35] proposed a neural network model that aggregated signals from external evidence articles that were retrieved with a search engine, the language of these articles and the trustworthiness of their sources. Karadzhov et al. [36] also used a search engine to retrieve documents relevant to the claim from which they built dense representations of the claim, the snippets and the related sentences which are then fed into an LSTM model.

Until recently, the majority of the proposed models were based on non-contextualized representations such as bag-of-words and non-contextual word embeddings. However, the effectiveness of context representation of a sentence using the Bidirectional Encoder Representations from Transformers (BERT) model led some researchers to apply it on the field of fake news detection [22,37]. Jwa et al. [22] used BERT to analyze the relationship between the headline and the body text and its impact on fake news detection, whereas Kaliyar et al. [37] combined different parallel blocks of the single-layer CNN with BERT. In our study, we also use a contextualized representation of the text that is based on BERT embeddings.

2.2. The role of users in disinformation

Bots and humans play an important role in the dissemination of fake news. Humans are not very effective in detecting fake news [9] because they tend to believe what confirms their prior beliefs or values [10] or perceptions that their peers share and believe [11]. On the other hand, humans have a significant contribution to fake news distribution. This was supported by a study from Vosoughi et al. [13] who showed that bots and humans accelerated the spread of true and false news at the same rate.

Different studies have tried to either profile the users that are involved in the different phases of fake news creation and propagation or to explore their impact on fake news detection. Some studies have employed survey based methodologies to profile users that tend to believe fake news. Pennycook and Rand [38] investigated the psychological profile of people that believed fake news by recruiting 1,606 participants for three online surveys. The study showed that people who believed in fake news were also accepting weak claims, were more willing to overclaim knowledge, and tended to get low scores in analytical thinking tests. Wolverton and Stevens [39] showed in a quantitative research that individuals with specific personality traits (i.e., closed to experience or cautious, introverted, disagreeable or unsympathetic, unconscientious or undirected and emotionally stable) are better in identifying disinformation.

Other studies proposed computational approaches to profile users who shared fake news using information from their social network accounts and posts. Shu et al. [16] analyzed different explicit and implicit characteristics of the users (e.g., registration time, location, age) that shared fake news and found that some of the most useful characteristics are registration time and whether the user was verified or not. Different to this study, we focus on a wide variety of psycholinguistic characteristics of spreaders and checkers as we are interested in studying their impact on their discrimination.

Rangel et al. [40] organized an evaluation task on profiling fake news spreaders. In order to create the collection, they first collected tweets that were relevant to news published in fact-checking websites. Then, they collected the tweets of users that had previously shared fake news. The dataset contained tweets in both English and Spanish. A wide range of features (e.g., word embeddings, n-gram, emotions, personality traits) and learning approaches (e.g., Support Vector Machines, Logistic Regression, CNN, LSTM) were explored by the participants to address the task of fake news spreaders detection. The highest performances were achieved by the approaches that used n-grams [41,42]. In particular, Pizarro [41] addressed the task by training a SVM with combinations of character and n-grams, whereas Buda and Bolonyai [42] used a Logistic Regression ensemble of five sub-models. Four models (LR, SVM, Random Forest, XGBoost) were based on n-grams, while the last one, an XGBoost utilized features that modeled various textual descriptive statistics such as the average length of the tweets.

Other studies focused on users that aimed to raise awareness by sharing posts from fact-checking websites. Vo and Lee [43] analyzed linguistic characteristics of fact-checking tweets and found that these tweets contain more formal language and less swear words. Giachanou et al. [17] focused on conspiracy theories that are a particular type of disinformation and on the detection of users that tend to share them, known as conspiracy propagators. In their study, they performed a comparative analysis over various profiles, psychological and linguistic characteristics of conspiracy and anti-conspiracy propagators. Their model that was based on a CNN network trained with word embeddings and psycho-linguistic characteristics, managed to effectively differentiate conspiracy and anti-conspiracy propagators. Similar to their study, we also explore different linguistic characteristics. However, in our study we are interested in exploring the differences between users that spread fake news and those that debunk them and explore their effectiveness on the task of spreaders and checkers discrimination.

3. Methodology

In this section, we first outline the process that we followed to create the collection and then we present the CheckerOrSpreader model.

3.1. Collection

Different collections have been developed for the field of disinformation detection, with the majority of them focusing on the credibility at article or post level [44,45]. The most common process to create a collection of fake news is to leverage the articles that are posted by one or more fact-checking websites. The collection can be based only on those articles or on their expanded version by including relevant posts shared on Twitter or other social networks. There are also some user-level collections, including the one developed for the CLEF evaluation task that contains tweets of users that have and have not spread fake news [40]. However, this collection does not contain tweets by fact-checkers. Vo and Lee [43] focused on fact-checking but they collected fact check tweets and not previous tweets posted by users. To the best of our knowledge, there is no collection that we can use for the task of differentiating users as checkers or spreaders. Therefore, we decided to build our own collection.[4]

Similar to previous research, we started with collecting articles from a fact-checking website, the Lead Stories. From this initial step, we collected 915 titles of articles that had been already labeled as fake by experts. Then, we removed stopwords from the headlines, which are then used as queries to search for relevant tweets. Fig. 1 shows the pipeline that we followed to create the collection.

To retrieve the relevant tweets, we used Twitter API with which we managed to collect 18,670 tweets that were referring to the lead stories articles. Due to the fact that some of the articles were discussed by more users compared to others, some of them have a higher number of posts. Table 1 shows examples of articles for which we collected the highest and lowest number of tweets. From this table, we observe that the most popular article was about a medical topic for which there were 1,448 relevant tweets. In addition, Fig. 2 shows the number of collected tweets per article. We observe that the frequencies follow a heavy-tailed distribution since a lot of tweets were posted for few articles and very few tweets for a lot of articles.

The next step in the process is to determine whether a tweet debunked the original article by claiming its falseness (fact check tweet) or shared the article claiming its truthfulness (spreading tweet). The fact-checking tweets contain specific patterns that can

[4] The collection and the code will be available upon acceptance.

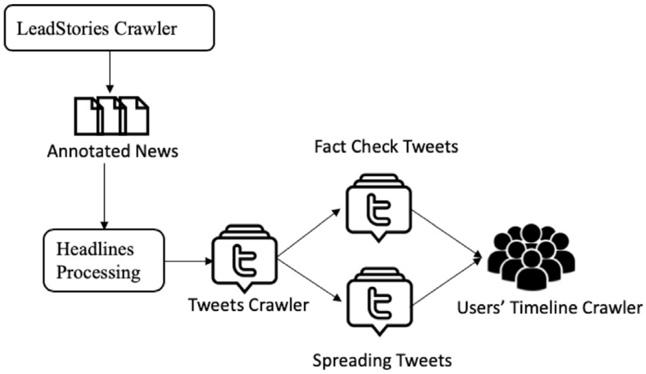


Fig. 1. Pipeline for the creation of the collection.

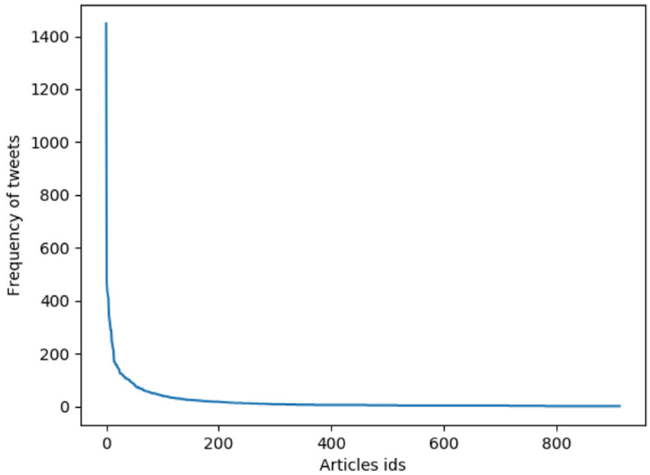


Fig. 2. Frequency distribution regarding of the number of tweets per article.

Table 1
Titles of the articles with the highest and lowest number of tweets.

Titles of articles with the highest number of tweets	Titles of articles with the lowest number of tweets
1. Doctors Who Discovered Cancer Enzymes In Vaccines NOT All Found Murdered	1. Make-A-Wish Did NOT Send Terminally Ill Spider-Man To Healthy Kid
2. Sugar Is NOT 8 Times More Addictive Than Cocaine	2. Man Did NOT Sue Radio Station For Playing Despacito 800 Times A Day
3. George H.W. Bush Did NOT Die at 93	3. Man-Eating Shark NOT Spotted In Ohio River
4. NO Alien Invasion This Is NOT Real	4. FBI DID NOT Classify President Obama As A Domestic Terrorist
5. First Bee Has NOT *Just* Been Added to Endangered Species List	5. Will Smith IS NOT Leaving America With His Family Never To Come Back

be used to facilitate the annotation process. According to those manually identified patterns, the tweets that mentioned any of the terms {hoax, fake, false, fact check, snopes, politifact leadstories, lead stories} were annotated as a fact check tweet, otherwise it was labeled as spreading tweet.

Fig. 3 shows an example of an article debunked as fake together with fact check and spreading tweets. We notice that in the fact check tweets we have terms such as *fake*, *false* and *fact check*, whereas the spreading tweets are re-posts of the original article. In addition, to explore the effectiveness of the annotation process, we manually checked 500 tweets, and we did not find any cases of misclassification.

After the annotation of the tweets, it is possible to also annotate the authors of the tweets as *checkers* or *spreaders* based on the number of fact check and spreading tweets they posted. In the case of users having both fact check and spreading tweets, we classified these users in the category for which they have the larger number of tweets. Finally, we collected the timeline of tweets that the authors have posted to create our collection. In total, our collection contains tweets posted by 2,357 users, of which 454 are checkers and 1504 spreaders.

3.2. The CheckerOrSpreader model

CheckerOrSpreader is based on a CNN. The architecture of the CheckerOrSpreader system is depicted in Fig. 4. CheckerOrSpreader consists of two different components, the embeddings layer and the user's psycholinguistic component. For the embeddings

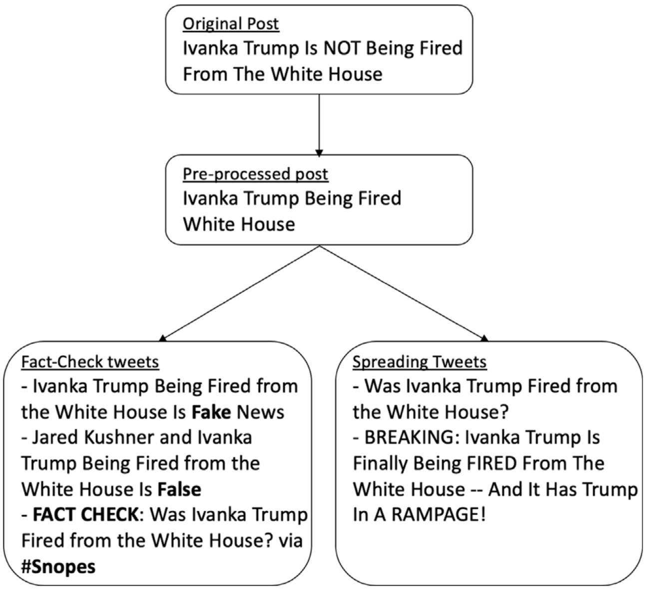


Fig. 3. Examples of fact check and spreading tweets.

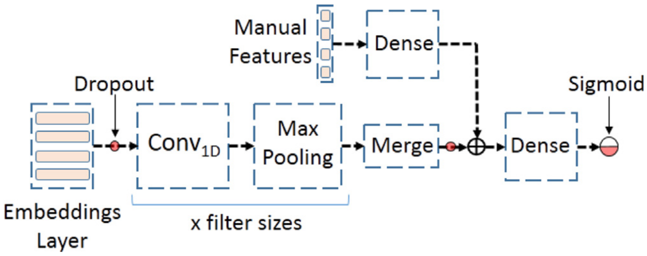


Fig. 4. Architecture of the CheckerOrSpreader model.

layer, we use BERT [46]. In particular, we represent all the tweets of a user using BERT embeddings and then we apply mean operation to calculate an average vector per user. The psycholinguistic component represents the psychometric and linguistic style patterns that were derived or inferred from the textual content of the posts. These include sentiment, emotions, linguistic patterns, personality traits, communication style and readability.

3.2.1. *Contextualized word embeddings*

The textual information is the most important element for the detection of fake news and fake news spreaders. It has been shown to be useful for a wide range of text classification tasks, from reputation analysis to irony detection and from author profiling to credibility detection [27,42,47,48]. Previous approaches of fake news spreaders detection have used representations such as bag-of-words and word embeddings [16,41,42]. These representations cannot capture contextual information. On the contrary, we propose to use a more sophisticated representation that is based on BERT [46]. BERT has been already applied in the field of fake news detection and has showed that can be useful for the task [33,37]. BERT applies the bidirectional training of a Transformer, an attention mechanism that learns contextual relations among words in a text. The Transformer includes an encoder and a decoder as two separate mechanisms. The encoder reads the text input and the decoder outputs a prediction for the task. Different to previous systems that looked at a text sequence either from left to right or combined left-to-right and right-to-left training, the Transformer encoder reads the entire sequence of words at once. This characteristic allows the model to learn the context of a word based on all of its surroundings (left and right of the word).

In particular, we use the pretrained BERT-Base available on the TensorFlow Hub.[5] This version allows an input text that has a maximum length of 512 that is much smaller compared to the concatenated tweets. To overcome this issue, we first extracted the embeddings of each individual post, and then we averaged the embeddings computed for each post in order to obtain a single representation.

[5] See https://www.tensorflow.org/hub.

3.2.2. Linguistic patterns

For the linguistic patterns, we employ LIWC [49], a dictionary-based software for mapping text to 73 psychologically-meaningful linguistic categories.[6] In particular, we extract personal pronouns (I, we, you, she/he, they), personal concerns (work, leisure, home, money, religion, death), cognitive processes (insight, causation, discrepancy, tentative, certainty, differentiation), informal language (swear, netspeak, assent, nonfluencies, fillers), perceptual processes (see, hear, feel), affective processes (positive emotion, negative emotion, anxiety, anger, sadness) and punctuation marks (question marks, exclamation marks).

3.2.3. Emotions

Some studies have showed that fake news triggers different emotions at a varied intensity compared to real news. To this end, we explore whether the tweets written by spreaders express different emotions compared to the tweets posted by checkers. We calculate scores for eight emotions: *anger, anticipation, disgust, fear, joy, sadness, surprise, trust* that are the core emotions according to Plutchik model [50]. To calculate the scores we follow a lexicon-based approach by utilizing the frequencies of the emotional words in the NRC emotions lexicon [51]. NRC lexicon is a list of about 14,000 English words which are manually labeled regarding different emotions and sentiments.

3.2.4. Sentiment

Sentiment expressed within news articles is likely to be different within fake news and true news [52]. Based on that, we assume that spreaders and checkers use sentiment words with a different frequency in their posts. To calculate the positive and negative sentiment in users' tweets we use a lexicon-based approach and the NRC emotions lexicon [51].

3.2.5. Personality traits

Previous research has showed that individuals with specific personality traits tend to better understand which articles are fake [39]. In addition, Shu et al. [16] used an unsupervised personality prediction tool to calculate personality scores of users based on their posts and showed that users that spread fake news have lower extraversion and agreeableness scores compared to those that spread real news.

Similar to [16] we also follow a computational approach to calculate personality scores based on the users' posts. The Five-Factor Model (FFM) [53], also called the Big Five, is the most popular methodology used in automatic personality research [54,55]. In essence, it defines five basic *factors* or *dimensions* of personality. These factors are:

- *openness to experience* (unconventional, insightful, imaginative)
- *conscientiousness* (organized, self-disciplined, ordered)
- *agreeableness* (cooperative, friendly, empathetic)
- *extraversion* (cheerful, sociable, assertive)
- *neuroticism* (anxious, sad, insecure)

Each of the five factors presents a *positive* and a complementary *negative* dimension. For instance, the complementary aspect to neuroticism is defined as *emotional stability*. Each individual can have a combination of these dimensions at a time. To obtain the personality scores, we followed the approach developed by Neuman and Cohen [56]. For their approach, they first proposed the construction of a set of vectors using a small group of adjectives, which according to theoretical and/or empirical knowledge, encode the essence of personality traits and personality disorders. Using context-free word embeddings, they measured the semantic similarity between these vectors and the text written by different individuals. The estimated similarity scores, allowed to quantify the degree in which a particular personality trait or disorder was evident in the text.

3.2.6. Communication style

The intention of written text can be very useful for the detection of fact checkers and fake news spreaders. To extract the intention of the posts of users we use Symanto API.[7] We calculate scores for *action-seeking* (calling for action or attention), *fact-oriented* (discussing about factual information), *information-seeking* (asking questions, seeking advice), *self-revealing* (sharing one's own opinion and experience). In addition, we calculate *emotional* and *rational* scores that represent whether the text is rational or not.

3.2.7. Readability

Readability assessment aims to estimate the complexity of a text in order to predict if a reader of a certain level of literacy will understand it. Previous studies have explored a wide range of readability features and have showed their impact on fake news detection [19,57,58]. Pérez-Rosas et al. [57] extracted 26 different readability features which achieved the best results in a dataset of fake news collected by crowdsourcing, whereas Santos et al. [58] showed the effectiveness of the readability features on a Brazilian Portuguese fake news dataset.

Based on the findings of the previous studies, we are also interested in seeing if readability can have an impact on spreaders and checkers differentiation. We use the following three scores: Flesch Kincaid Reading Ease (FKRE), Simple Measure of Gobbledygook (SMOG), Automated Readability Index (ARI).

[6] For a comprehensive list of LIWC categories see: http://hdl.handle.net/2152/31333.

[7] See https://developers.symanto.net/.

Table 2
Parameter optimization for the different tested systems.

	filters sizes	# of filters	activation	optimizer	epochs
LSTM-embeddings	64 (lstm units)	-	tanh	rmsprop	11
CNN-embeddings	4	32	relu	adadelta	12
CheckerOrSpreader	3,5	8	selu	adadelta	17

4. Experimental setup

In this section, we present the experimental setup that we followed for our experiments, including the experimental settings and the evaluation process.

4.1. Experimental settings

For our experiments, we use 25% of our corpus of users for validation, 15% for test and the remaining 60% for training. For two of the baselines (LSTM-embeddings and CNN-embeddings), we use the 300-dimensional pre-trained GloVe embeddings [59]. Table 2 shows the optimization parameters for LSTM-embeddings, CNN-embeddings and CheckerOrSpreader approaches. For the textual representation, we use the pretrained BERT-Base available on TensorFlow Hub. In particular, we extracted the embeddings of each individual post, and then we averaged the embeddings computed for each post to get a single representation.

4.2. Evaluation

For the evaluation, we present Precision, Recall and macro-F1 score. For some baselines, where classical classifiers are used, we tested Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Gaussian Naïve Bayes (GNB), and we reported the result of the one that achieved the highest F1-score on the validation set. We use the following baselines to compare our results:

- *Random Forest (RF)*: We use Random Forest trained on the different linguistic and personality scores features. In particular, we tried sentiment, LIWC, personality traits, readability and communication style.
- *Emotion-GNB*: We train a GNB classifier with emotions extracted from the text of the tweets. We use NRC emotions lexicon [51] and we extracted anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.
- *BERT-LR*: For the BERT-LR baseline, we estimate the average vector embeddings of the final tweets represented using the pre-trained BERT embeddings which are then fed into a LR classifier.
- *USE-LR*: This baseline is based on Universal Sentence Encoder (USE) [60]. For the USE-LR baseline, we represent the final concatenated documents (tweets) using USE embeddings[8] which are then fed into a LR classifier.
- *LSTM-embeddings*: is based on a LSTM network with Glove pre-trained word embeddings for word representation.
- *CNN-embeddings*: is a CNN with Glove pre-trained word embeddings for word representation.

5. Results and analysis

In this section we first present the performance results of the methodology and then we analyze the differences and importance of the different psycholinguistic characteristics.

5.1. Results

Table 3 shows the results of our experiments when one group of features was used. From the results we observe that the best results were achieved for the *BERT-LR* and the *Personality-RF* combinations. We observe that emotion and sentiment achieve a lower performance compared to the *BERT-LR*. Emotions have been shown to be very useful in the field of fake news detection [27,28,61]. One possible explanation why emotions were not useful is that differently from previous studies, we extracted the overall scores of every emotion expressed in the posts of the checkers and spreaders and not emotions at the fake posts level. Also, from the results we observe that the readability and the LIWC features were not very helpful when used alone.

Table 4 shows the results of the CheckerOrSpreader model and the baselines. We observe that CNN performs similarly with LSTM when they are trained only using word embeddings. Also, we observe that CheckerOrSpreader managed to achieve the best performance. In particular, CheckerOrSpreader managed to improve the performance by 22% in terms of F1-metric compared to the CNN-embeddings baseline. Our model also outperforms the second best result that is achieved by *BERT-LR* by 7.02% in terms of F1-metric. This result suggests that the different psycholinguistic characteristics that we extracted from the users' posts can be useful for the discrimination of spreaders and checkers.

[8] See https://tfhub.dev/google/universal-sentence-encoder-large/3.

Table 3
Performance of the different systems when one group of features is used.

	Precision	Recall	F1-score
BERT-LR	0.58	0.57	0.57
Emotion-GNB	0.66	0.55	0.54
Sentiment-RF	0.59	0.55	0.54
LIWC-RF	0.58	0.53	0.51
Personality-RF	0.62	0.57	0.57
Readability-RF	0.58	0.53	0.51
CommunicationStyle-RF	0.62	0.55	0.54

Table 4
Performance of the different systems on the fact checkers detection task.

	Precision	Recall	F1-score
BERT-LR	0.58	0.57	0.57
USE-LR	0.79	0.86	0.53
LSTM-embeddings	0.56	0.51	0.49
CNN-embeddings	0.53	0.51	0.50
CheckerOrSpreader	0.62	0.60	0.61

5.2. Analysis

In this section, we perform further analysis with the aim to get a better understanding of the impact of the different psycholinguistic information on the discrimination of fake news spreaders and fact checkers.

5.2.1. Differences of the psycholinguistic characteristics

Here, we focus on the second research question: *RQ2. Which psycholinguistic variables show a statistically significant univariate difference between fake news spreaders and fact checkers?* To address this question, we analyze the distribution of each of psycholinguistic features in the spreaders and checkers tweets. Fig. 5 shows the distributions and means of different psycholinguistic features that we explored in our study.

One observation is that spreaders tend to have high amount of informal language. In particular, spreaders tend to use more netspeak (e.g., lol, omg) terminology ($\mu_{checkers}$ = 735.87 versus $\mu_{spreaders}$ = 869.81) and more swear words ($\mu_{checkers}$ = 12.08 versus $\mu_{spreaders}$ = 15.62) compared to checkers. This result is consistent to Vo et al. [43] who found that at a tweet level, spreading tweets contain more informal language compared to the fact-checking tweets.

Another interesting observation is that checkers use more words that show insight ($\mu_{checkers}$ = 105.17 versus $\mu_{spreaders}$ = 88.78) and causation ($\mu_{checkers}$ = 101.65 versus $\mu_{spreaders}$ = 79.75). A possible explanation is that they tend to explain more the information they share. A similar finding was also found by [17] who found that conspiracy propagators use less causation language compared to those that do not.

Another observation from the analysis of the different features is that checkers use more information-seeking language compared to the spreaders ($\mu_{checkers}$ = 0.063 versus $\mu_{spreaders}$ = 0.046), and they contain more rational language ($\mu_{checkers}$ = 0.626 versus $\mu_{spreaders}$ = 0.587).

In addition, we calculate the t-statistics differences of the features to see which are the psycholinguistic features for which there are statistically significant differences when they are used by spreaders and checkers. Fig. 6 shows the results after calculating those values. From the figure, we first observe that checkers tend to use more positive language compared to spreaders in a statistically significant way. In addition, checkers use more words that show causation (e.g., because) and insight (e.g., think). Also, their posts contain more information-seeking language. We also observe that they significantly use more terms that express trust, anticipation, surprise, anger and fear.

On the other hand, spreaders tend to use a higher number of informal netspeak and swear words. Also, they use a higher number of self-revealing language that represents the tendency to share your own experience and opinion. Regarding the personality features, we could not find any statistically significant differences in the means of checkers and spreaders. Moreover, we observed that personality traits correlate very strongly with each other, with most correlations among ostensibly different personality measures close to 0.99. We believe that one reason for this is the fact the approach that we used has been utilized and evaluated on less informal texts (e.g., stream-of-consciousness essays) than those found in social media. Therefore, a possible explanation is that the social media posts do not contain (or contain very few) terms that are semantically similar to the original adjectives used to characterize the Big Five traits and personality disorders. This finding implies that the use of the out-of-the box tools that have been extensively used in research should be made with more attention and that post analysis of this information can increase the confidence in their calculations and in using those tools.

5.2.2. Importance of the psycholinguistic characteristics

In this section, we aim to address the question of *RQ3. What are the top-N psycholinguistic characteristics that discriminate fake news spreaders and fact checkers in a multivariable prediction?*. To address this question, we compute the relative importance of each

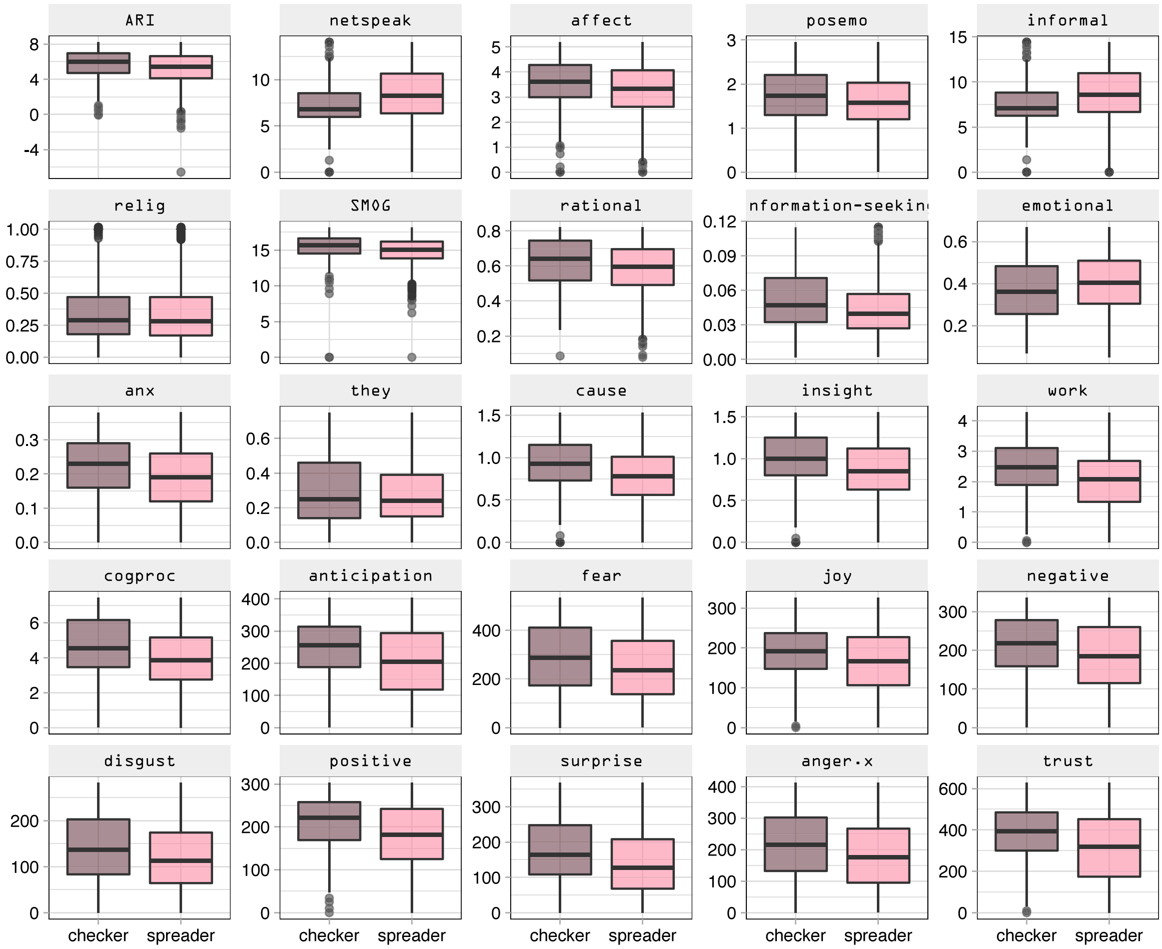


Fig. 5. Distributions of different psycholinguistic features of spreaders and checkers.

psycholinguistic attribute with respect to the target class. In particular, we follow two different but complementary approaches. Linear classification algorithms, such as SVM, attempt to find a separating line (or hyperplane) between the samples, generally, of two classes. In this way, the class assigned to a subject is the result of the weighted sum of the input features. We first analyze the weights obtained in the training stage of a SVM as a way to better understand the contribution that each feature has when predicting the class of an unseen sample.

Fig. 7 shows the contribution of each feature when predicting the class of an unseen sample for all the features combined. From this analysis, we observe that the highest contribution for the spreader class was obtained by the informal language, causality and question marks. On the other hand, the highest contribution to the prediction of the checker class was made by swear words and personal pronouns.

An alternative way to study the contribution that each feature has in the classification process is to analyze their expected mutual information (MI). In essence, MI measures how much information an attribute contains of the target class. If an attribute's distribution is the same in the class as it is in the dataset as a whole, then MI is equal to zero. MI reaches its maximum value if the attribute is a perfect indicator for class membership. To this end, we analyze the MI computed for each feature.

Fig. 8 shows the mutual information of different psycholinguistic features. Fig. 8(a) and 8(b) show the mutual information obtained for the sentiment and emotions features respectively. We observe that the highest MI score for sentiment was obtained by positive language, whereas trust for the emotions. In particular, trust, joy and anticipation are those that achieved the three top MI scores. From Fig. 8(c) we observe that the top features were the information-seeking and fact-oriented language. From Fig. 8(d) we observe that the highest MI is obtained by the informal language. This is in compliance to the analysis that we have done in the previous section. Also, we see that the second and third highest MI scores are obtained by the *work* related terms and question marks. On the contrary, we observe that personal pronouns and exclamation marks did not obtain high MI scores.

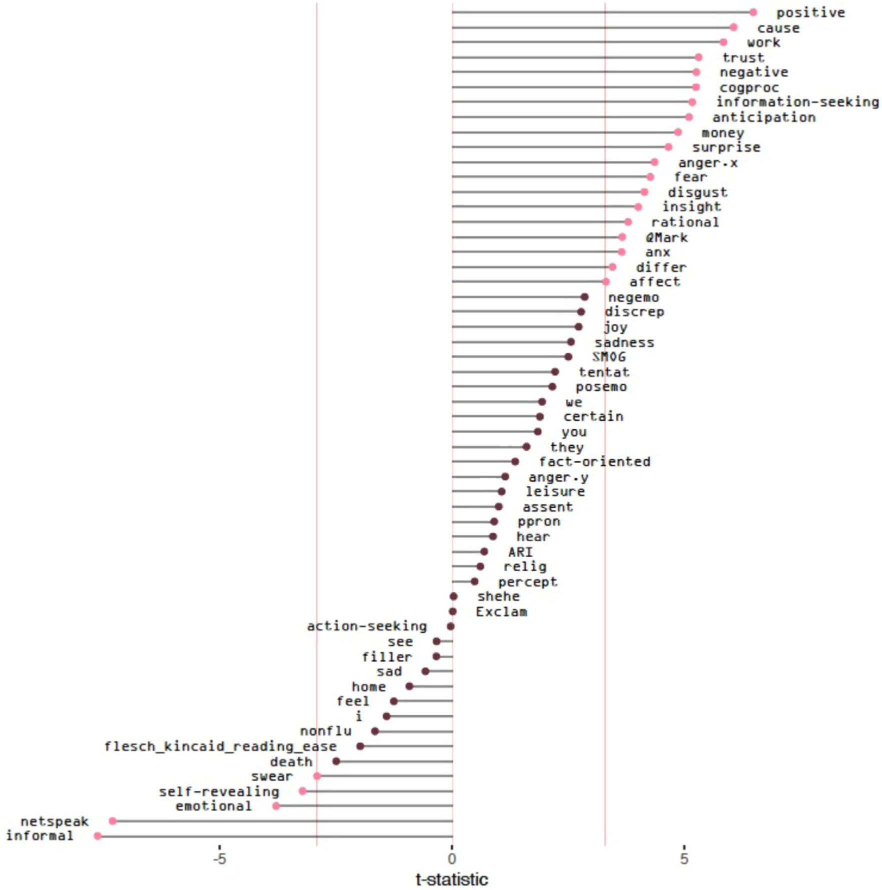


Fig. 6. T-statistics for mean differences with FDR-adjusted *p*-values of the different psycholinguistic features. *P*-values that are less than 0.01 are colored with pink. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Limitations and ethical concerns

Even if our study can provide valuable insights regarding the profile of spreaders and their automated detection, there are some limitations and ethical concerns that should be discussed. One limitation of our study is the use of an automated tool to infer the personality traits of the users based on the tweets that they have posted. Even if this tool has been shown to achieve good prediction performance on other data sets, it is still prone to errors similar to all the automated tools. Also, another limitation with this tool and other similar ones is that they have been evaluated on regular text and not on social media texts. That means that some of the predictions regarding the personality traits that were inferred from the tweets might not be accurate. Also, the fact that our analysis showed that the personality traits give nearly identical scores (up to a constant linear transformation) shows the limitation of the tool to make accurate predictions on social media text. Unfortunately, it is not possible to confirm any of those claims because we cannot evaluate the performance of this tool on our collection given that we do not have ground truth data regarding the users' personality traits. An alternative way to obtain information regarding the personality traits would be to contact these users and ask them to fill one of the standard questionnaires (e.g., IPIP questionnaire [62]) that have been evaluated based on several psychological studies and tend to have more precise results. However, this is beyond the scope of the current study.

Our study has also some ethical concerns given that it is focused on a user-level classification. Here, we want to highlight that a system that can differentiate between spreaders and checkers should by no means be used to stigmatize a user that has shared fake news in the past. Instead, it should be used to facilitate research in fake news detection, and this can be done in different ways. For example, the system could be used as a supportive tool to alert spreaders in order to raise awareness especially of those that unintentionally share fake news. Another way could be to recognize checkers and make their fact-checking tweets more visible to other users. We also want to highlight that a system that differentiates users between spreaders and checkers requires to consider ethics at all steps.

Finally, our study has also some ethical concerns regarding the collection and the release of the data. First, we plan to make this collection available only for research purposes. To protect the privacy of users, we plan to publish the data anonymized. Also, we plan to use neutral annotation labels regarding the two classes (i.e., 0 and 1 instead of checker and spreader) since we do not

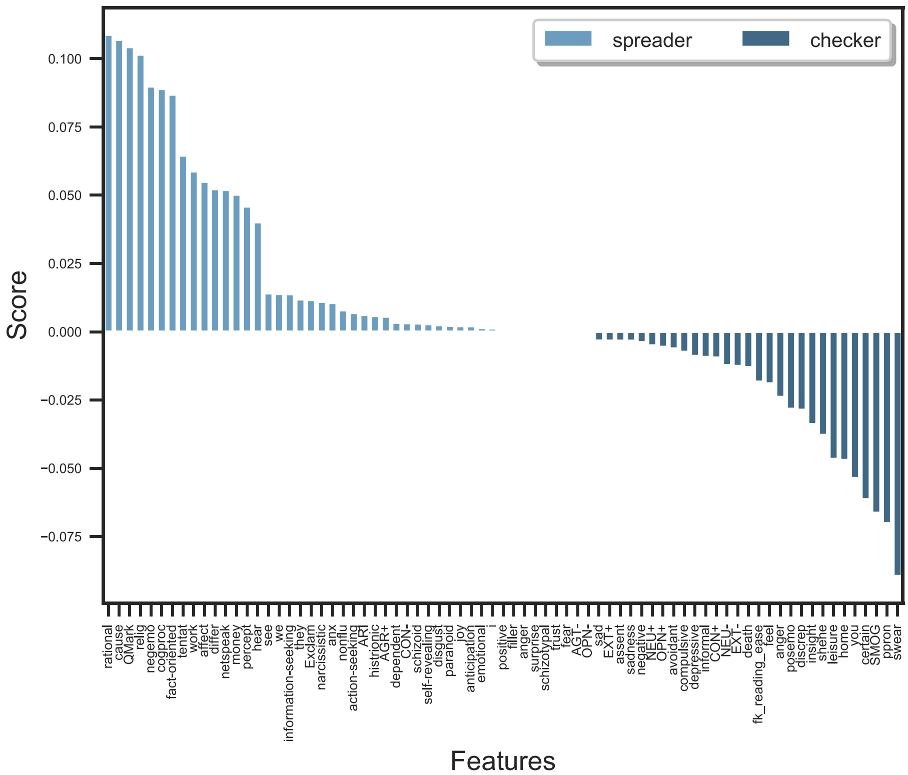


Fig. 7. Contribution of feature when predicting the class of an unseen sample for all the features combined.

want to stigmatize specific users. Future researchers that want to use the collection will not have access to the information related to which class each label refers to. Finally, we will not make available the labels at a post level since this information can reveal the information regarding the annotation labels at a user level.

7. Conclusions and future work

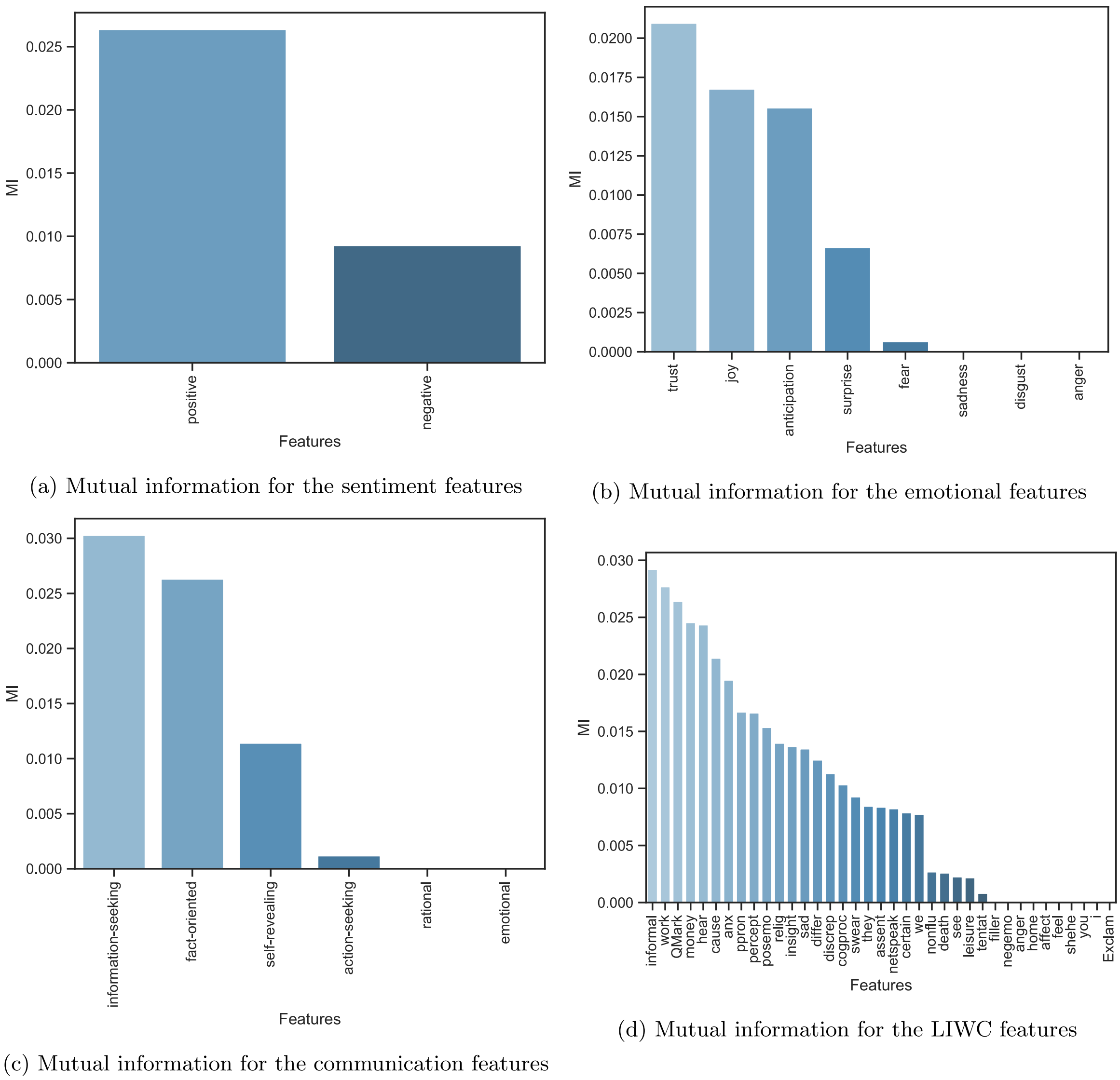
In this paper, we focused on the problem of differentiating between users that tend to share fake news (spreaders) and those that tend to check the factuality of articles (checkers). To address the problem, we proposed CheckerOrSpreader, a model based on a CNN and which incorporated various psycholinguistic characteristics that were extracted or inferred from users' posts. In particular our model incorporated sentiment, emotions, linguistic features, personality traits, readability and communication style extracted from users' posts. For the textual representation, we proposed to use BERT embeddings.

In addition, we presented a collection that can be used for the problem of differentiating between spreaders and checkers. Experimental results showed that the CheckerOrSpreader model managed to achieve the best performance compared to the baselines, a finding that suggested that the psycholinguistic characteristics could be useful for the task.

Also, with the aim to get a better understanding of the features we estimated the statistical differences of the various psycholinguistic characteristics of checkers and spreaders. Our analysis suggested that spreaders tend to use more informal language including netspeak and swear words compared to the checkers. On the contrary, checkers used a higher number of positive words. Also, they tended to use a higher number of words that refer to causality and work. Checkers also used more words that show trust, anticipation and anger. Our analysis also showed that some of the most important communication style features according to mutual information were information-seeking language, whereas regarding the emotional features the top feature was trust. Those results could be used to design a classifier that would be able to distinguish spreaders from checkers.

Finally, our analysis showed that the estimated personality traits features were highly correlated to each other. This finding implies that the method that we used to estimate the personality traits did not work as intended for the specific collection. One reason is that our collection contained tweets, whereas the methodology that we used was evaluated on regular text. We believe that this is an important finding because it shows that the automatic way of extracting features is prone to errors and researchers should pay more attention when they use them.

In the future, we plan to further investigate and explore ways of addressing the current limitations of our study. In particular, we are interested to better understand the reasons why the personality traits were correlated among them and if the vocabulary of the posts had an impact on that. In addition, we plan to investigate how the various psycholinguistic characteristic that can be extracted from users' posts and the information regarding whether a user is a potential spreader or checker can be incorporated into the systems that detect fake news.



(a) Mutual information for the sentiment features

(b) Mutual information for the emotional features

(c) Mutual information for the communication features

(d) Mutual information for the LIWC features

Fig. 8. Mutual information for different psycholinguistic features.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The works of Anastasia Giachanou and Daniel Oberski were funded by the Dutch Research Council (grant VI.Vidi.195.152). The work of Paolo Rosso was in the framework of the XAI-DisInfodemics project on eXplainable AI for disinformation and conspiracy detection during infodemics (PLEC2021-007681), funded by the Spanish Ministry of Science and Innovation, as well as IBERIFIER, the Iberian Digital Media Research and Fact-Checking Hub funded by the European Digital Media Observatory (2020-EU-IA0252).

References

[1] T. Lee, The global rise of "fake news" and the threat to democratic elections in the USA, Public Adm. Policy (2019).
[2] A. Bovet, H.A. Makse, Influence of fake news in Twitter during the 2016 US presidential election, Nature Commun. 10 (1) (2019) 1–14.
[3] F. Pierri, A. Artoni, S. Ceri, Investigating Italian disinformation spreading on Twitter in the context of 2019 European elections, PLoS One 15 (1) (2020) e0227821.

[4] M.T. Bastos, D. Mercea, The brexit botnet and user-generated hyperpartisan news, Soc. Sci. Comput. Rev. 37 (1) (2019) 38–54.
[5] G. Pennycook, J. McPhetres, Y. Zhang, J.G. Lu, D.G. Rand, Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention, Psychol. Sci. 31 (7) (2020) 770–780.
[6] S.O. Oyeyemi, E. Gabarron, R. Wynn, Ebola, Twitter, and misinformation: A dangerous combination? BMJ Clin. Res. 349 (2014).
[7] D. DiFranzo, K. Gloria-Garcia, Filter bubbles and fake news, ACM Crossroads 23 (3) (2017) 32–35.
[8] World Health Organization, et al., Novel Coronavirus (2019-nCoV): Situation Report, 3, World Health Organization, 2020.
[9] V.L. Rubin, On deception and deception detection: Content analysis of computer-mediated stated beliefs, Proc. Am. Soc. Inf. Sci. Technol. 47 (1) (2010) 1–10.
[10] P. Moravec, R. Minas, A.R. Dennis, Fake news on social media: People believe what they want to believe when it makes no sense at all, Kelley Sch. Bus. Res. Pap. (18–87) (2018).
[11] Q. Wang, X. Yang, W. Xi, Effects of group arguments on rumor belief and transmission in online communities: An information cascade and group polarization perspective, Inf. Manage. 55 (4) (2018) 441–449.
[12] E. Ferrara, O. Varol, C. Davis, F. Menczer, A. Flammini, The rise of social bots, Commun. ACM 59 (7) (2016) 96–104.
[13] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, Science 359 (6380) (2018) 1146–1151.
[14] M.V. Bronstein, G. Pennycook, A. Bear, D.G. Rand, T.D. Cannon, Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking, J. Appl. Res. Mem. Cogn. 8 (1) (2019) 108–117.
[15] M.L. Stanley, N. Barr, K. Peters, P. Seli, Analytic-thinking predicts hoax beliefs and helping behaviors in response to the COVID-19 pandemic, Think. Reason. (2020) 1–14.
[16] K. Shu, S. Wang, H. Liu, Understanding user profiles on social media for fake news detection, in: Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval, MIPR '18, 2018, pp. 430–435.
[17] A. Giachanou, B. Ghanem, P. Rosso, Detection of conspiracy propagators using psycho-linguistic characteristics, J. Inf. Sci. (2021).
[18] A. Giachanou, E.A. Ríssola, B. Ghanem, F. Crestani, P. Rosso, The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers, in: Natural Language Processing and Information Systems, in: NLDB '20, Springer International Publishing, 2020, pp. 181–192.
[19] X. Zhou, A. Jain, V.V. Phoha, R. Zafarani, Fake news early detection: A theory-driven model, Digit. Threat.: Res. Pract. 1 (2) (2020).
[20] A. Giachanou, G. Zhang, P. Rosso, Multimodal fake news detection with textual, visual and semantic information, in: Text, Speech, and Dialogue, in: TSD '20, Springer, 2020, pp. 30–38.
[21] H. Rashkin, E. Choi, J.Y. Jang, S. Volkova, Y. Choi, 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, Association for Computational Linguistics, 2017, pp. 2931–2937.
[22] H. Jwa, D. Oh, K. Park, J.M. Kang, H. Lim, Exbake: Automatic fake news detection model based on bidirectional encoder representations from transformers (BERT), Appl. Sci. 9 (19) (2019) 4062.
[23] X. Zhou, R. Zafarani, A survey of fake news: Fundamental theories, detection methods, and opportunities, ACM Comput. Surv. 53 (5) (2020) 1–40.
[24] G. Ruffo, A. Semeraro, A. Giachanou, P. Rosso, Surveying the research on fake news in social media: a tale of networks and language, 2021.
[25] S. Volkova, K. Shaffer, J.Y. Jang, N. Hodas, 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 (Vol. 2: Short Papers), Association for Computational Linguistics, 2017, pp. 647–653.
[26] M. Zuckerman, B.M. DePaulo, R. Rosenthal, Verbal and nonverbal communication of deception, in: Advances in Experimental Social Psychology, Vol. 14, Elsevier, 1981, pp. 1–59.
[27] A. Giachanou, P. Rosso, F. Crestani, Leveraging emotional signals for credibility detection, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '19, 2019, pp. 877–880.
[28] B. Ghanem, P. Rosso, F. Rangel, An emotional analysis of false information in social media and news articles, ACM Trans. Internet Technol. (TOIT) 20 (2) (2020) 1–18.
[29] C. Guo, J. Cao, X. Zhang, K. Shu, H. Liu, DEAN: Learning dual emotion for fake news detection on social media, 2019, arXiv preprint arXiv:1903.01728.
[30] S. Singhal, R.R. Shah, T. Chakraborty, P. Kumaraguru, S. Satoh, SpotFake: A multi-modal framework for fake news detection, in: 2019 IEEE 5th International Conference on Multimedia Big Data, in: BigMM '19, IEEE, 2019, pp. 39–47.
[31] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, J. Gao, EANN: Event adversarial neural networks for multi-modal fake news detection, in: KDD'18, 2018, pp. 849–857.
[32] D. Khattar, J.S. Goud, M. Gupta, V. Varma, MVAE: Multimodal variational autoencoder for fake news detection, in: WWW '19, 2019, pp. 2915–2921.
[33] A. Giachanou, G. Zhang, P. Rosso, Multimodal multi-image fake news detection, in: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA '20, 2020, pp. 647–654.
[34] D. Zlatkova, P. Nakov, I. Koychev, Fact-checking meets fauxtography: Verifying claims about images, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 2099–2108.
[35] K. Popat, S. Mukherjee, A. Yates, G. Weikum, DeClarE: Debunking fake news and false claims using evidence-aware deep learning, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 22–32.
[36] G. Karadzhov, P. Nakov, L. Màrquez, A. Barrón-Cedeño, I. Koychev, Fully automated fact checking using external sources, in: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, 2017, pp. 344–353.
[37] R.K. Kaliyar, A. Goswami, P. Narang, FakeBERT: Fake news detection in social media with a BERT-based deep learning approach, Multimedia Tools Appl. (2021).
[38] G. Pennycook, D. Rand, Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking, J. Pers. (2018).
[39] C. Wolverton, D. Stevens, The impact of personality in recognizing disinformation, Online Inf. Rev. (2019).
[40] F. Rangel, A. Giachanou, B. Ghanem, P. Rosso, Overview of the 8th author profiling task at PAN 2020: Profiling fake news spreaders on Twitter, in: CLEF 2020 Labs and Workshops, Notebook Papers, CEUR Workshop Proceedings, 2020.
[41] J. Pizarro, Using N-grams to detect fake news spreaders on Twitter, in: CLEF 2020 Labs and Workshops, Notebook Papers, 2020.
[42] J. Buda, F. Bolonyai, An ensemble model using N-grams and statistical features to identify fake news spreaders on Twitter, in: CLEF 2020 Labs and Workshops, Notebook Papers, 2020.
[43] N. Vo, K. Lee, Learning from fact-checkers: Analysis and generation of fact-checking language, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'19, 2019, pp. 335–344.
[44] W.Y. Wang, Liar, liar pants on fire: A new benchmark dataset for fake news detection, CoRR (2017) arXiv:1705.00648.
[45] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, H. Liu, FakeNewsNet: A data repository with news content, social context and dynamic information for studying fake news on social media, CoRR (2018) arXiv:1809.01286.
[46] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, 2018, arXiv preprint arXiv:1810.04805.
[47] A. Giachanou, J. Gonzalo, F. Crestani, Propagating sentiment signals for estimating reputation polarity, Inf. Process. Manage. 56 (6) (2019) 102079.
[48] D.I.H. Farías, V. Patti, P. Rosso, Irony detection in Twitter: The role of affective content, ACM Trans. Internet Technol. (TOIT) 16 (3) (2016) 1–24.
[49] J.W. Pennebaker, R.L. Boyd, K. Jordan, K. Blackburn, The Development and Psychometric Properties of LIWC 2015, Technical report, 2015.
[50] R. Plutchik, A general psychoevolutionary theory of emotion, in: Theories of Emotion, Elsevier, 1980, pp. 3–33.

[51] S.M. Mohammad, P.D. Turney, 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, 2010, pp. 26–34.
[52] A.S. Ebesu, M.D. Miller, Verbal and nonverbal behaviors as a function of deception type, J. Lang. Soc. Psychol. 13 (4) (1994) 418–442.
[53] O.P. John, S. Srivastava, The big-five trait taxonomy: History, measurement, and theoretical perspectives, in: Handbook of Personality: Theory and Research, 1999, pp. 102–138.
[54] Y. Neuman, Computational Personality Analysis: Introduction, Practical Applications and Novel Directions, 1st, Springer Publishing Company, Incorporated, 2016.
[55] E.A. Ríssola, S.A. Bahrainian, F. Crestani, Personality recognition in conversations using capsule neural networks, in: 2019 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2019, Thessaloniki, Greece, October 14-17, 2019, 2019, pp. 180–187.
[56] Y. Neuman, Y. Cohen, A vectorial semantics approach to personality assessment, Sci. Rep. 4 (1) (2014).
[57] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news, in: Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, 2018, pp. 3391–3401.
[58] R. Santos, G. Pedro, S. Leal, O. Vale, T. Pardo, K. Bontcheva, C. Scarton, Measuring the impact of readability features in fake news detection, in: Proceedings of the 12th Language Resources and Evaluation Conference, European Language Resources Association, 2020, pp. 1404–1413.
[59] J. Pennington, R. Socher, C. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP '14, 2014, pp. 1532–1543.
[60] D. Cer, Y. Yang, S.Y. Kong, N. Hua, N. Limtiaco, R.S. John, N. Constant, M. Guajardo-Céspedes, S. Yuan, C. Tar, Y.H. Sung, B. Strope, R. Kurzweil, Universal sentence encoder, CoRR (2018) arXiv:1803.11175.
[61] B. Ghanem, P. S., P. Rosso, F. Rangel, FakeFlow: Fake news detection by modeling the flow of affective information, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2021.
[62] L.R. Goldberg, A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models, Pers. Psychol. Eur. 7 (1) (1999) 7–28.

Anastasia Giachanou is a postdoc fellow at the Human Data Science group at Utrecht University. Before that, she was a postdoc at Pattern Recognition and Human Language Technology (PRHLT) research center the Universitat Politècnica de València (UPV) working on the topic of fake news detection on social media. She received her Ph.D. from the Faculty of Informatics at the University of Lugano, in Switzerland where she was a member of the Information Retrieval group working on the topic of tracking sentiment over time. She is particularly interested in research topics of Natural Language Processing, Social Media Analysis and Computational Social Science. Her current research interests include advancing social science with NLP, fake news detection, bot detection, fact checking, and sentiment analysis. Other research interests include temporal opinion mining, professional search, and user experience.

Bilal Ghanem received a B.Sc. degree in Computer Information Technology at Arab American University, Palestine, in 2014. Then he joined a Master's degree in Computer Science at the same university, where he received his degree in 2017. During his Master's degree, he worked as Software Engineer in a local company, and then he moved to work at Al-Quds Open University as Database Data Auditor. After that, he started his Ph.D. at Technical University of Valencia, Spain. Bilal's research interests focus on developing machine learning models for a variety of tasks that use Natural Language Processing, with a particular focus on Fake News detection and analysis.

Esteban A. Ríssola received his Ph.D. from the Faculty of Informatics, Università della Svizzera italiana (USI) in Lugano, Switzerland. His research focuses on studying and understanding how individuals' mental states and personality manifest through online social interactions and how such knowledge can be used to address various tasks, such as the early identification of mental disorders or personality assessment. Formerly, he was a member of the research and development supporting team at Center for Research, Teaching and Transferring on ICT (CIDETIC), Universidad Nacional de Luján (UNLu), Argentina. He holds a licentiate degree (B.Sc. + M.Sc.) in Information Systems from UNLu. During his master's thesis he focused on the development of inverted index management strategies for improving real-time search efficiency (e.g., for microblogging services).

Paolo Rosso is Full Professor at the Universitat Politècnica de València, Spain where he is also a member of the PRHLT research center. His research interests are focused on social media data analysis, mainly on author profiling, irony detection, fake news, and hate speech detection. Since 2009 he has been involved in the organization of PAN benchmark activities, mainly in the framework of CLEF, where he is also deputy steering committee chair for the conference, and FIRE evaluation forums, on plagiarism/text reuse detection and author profiling. He has been also coorganizer of shared tasks at SemEval on sentiment analysis of figurative language in Twitter (2015) and hate speech detection (2019), as well as at the Spanish and Italian evaluation forums of IberEval (now IberLEF) and Evalita. He has been PI of several national and international research projects funded by EC, U.S. Army Research Office, and Qatar National Research Fund. At the moment he is the PI of MISMIS-FAKEnHATE, a research project funded by the Spanish Ministry of Science and Innovation on Misinformation and Miscommunication in social media: FAKE news and HATE speech. He is associate editor at Information Processing & Management. He is co-author of 50+ articles in international journals and 400+ articles in conferences and workshops.

Fabio Crestani is a Full Professor at the Faculty of Informatics of the Università della Svizzera italiana (USI) in Lugano, Switzerland, since 2007. Previously, he was a professor at the University of Strathclyde (UK, 2000–06), a postdoc (1997–99) at the University of Glasgow (UK), the International Computer Science Institute in Berkeley (USA), and the Rutherford Appleton Laboratory (UK). Yet, it all started in 1992–96 when he was Assistant Professor at the University of Padua (Italy). He holds a degree in Statistics and Economics from the University of Padua (Italy) and a M.Sc. and a Ph.D. in Computing Science from the University of Glasgow (UK).

Daniel Oberski is a Full Professor at the department of Methodology and Statistics of Utrecht University in the Netherlands, currently working on the VIDI project from the Netherlands Organization for Scientific Research (NWO), "Advancing social science with valid measures derived from incidental data". His research focuses on the problem of measurement in the social sciences.