

The impact of emotional signals on credibility assessment

Anastasia Giachanou¹  | Paolo Rosso¹ | Fabio Crestani² 

¹Universitat Politècnica de València, Valencia, Spain

²Università della Svizzera italiana, Lugano, Switzerland

Correspondence

Anastasia Giachanou, Utrecht University, Utrecht, The Netherlands.
Email: a.giachanou@uu.nl

Present address

Anastasia Giachanou, Utrecht University, Utrecht, The Netherlands

Funding information

Generalitat Valenciana, Grant/Award Number: DeepPattern (PROMETEO/2019/121); Ministerio de Ciencia e Innovación, Grant/Award Number: PGC2018-096212-B-C31; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: P2TIP2_181441

Abstract

Fake news is considered one of the main threats of our society. The aim of fake news is usually to confuse readers and trigger intense emotions to them in an attempt to be spread through social networks. Even though recent studies have explored the effectiveness of different linguistic patterns for fake news detection, the role of emotional signals has not yet been explored. In this paper, we focus on extracting emotional signals from claims and evaluating their effectiveness on credibility assessment. First, we explore different methodologies for extracting the emotional signals that can be triggered to the users when they read a claim. Then, we present emoCred, a model that is based on a long-short term memory model that incorporates emotional signals extracted from the text of the claims to differentiate between credible and non-credible ones. In addition, we perform an analysis to understand which emotional signals and which terms are the most useful for the different credibility classes. We conduct extensive experiments and a thorough analysis on real-world datasets. Our results indicate the importance of incorporating emotional signals in the credibility assessment problem.

1 | INTRODUCTION

Fake news¹ is currently considered one of the greatest problems in our society. While fake news is not a new phenomenon, the rise of social media has facilitated their dissemination and has exacerbated the problem and its consequences. The extensive spread of false and inaccurate information has the potential for negative effects in society in connection to various domains such as politics, finance, and health. The impact of inaccurate information can be observed during the U.S. 2016 elections where, according to a number of commentators, the propagation of false and inaccurate information might have influenced the elections outcome.² In addition, in the final 3 months of the U.S. presidential campaign, the top 20 frequently-discussed false election stories

generated more engagements on Facebook than the top 20 most-discussed election stories posted by 19 major news websites.³ In addition, the false rumor that President Obama was injured after a terror attack on April 2013 caused a 143-point fall in the Dow Jones industrial average (Ferrara et al., 2016). In the health domain, anti-vaccine campaigns led to a decrease of Measles, Mumps, & Rubella vaccination rates⁴ causing in 2017 one of the worst measles outbreak in decades.

Understanding if a piece of information is accurate or not is a very challenging problem even for humans. The confusion is caused by the fact that fake news contain a mixture of false and true information (Shu et al., 2017). In addition, different types of fake news such as rumors and false claims usually cause different emotions and of different intensity to the readers in an attempt to be believed and

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Journal of the Association for Information Science and Technology* published by Wiley Periodicals LLC on behalf of Association for Information Science and Technology.

propagated in social networks (Vosoughi et al., 2018). To help humans detect inaccurate information, several fact-checking websites (e.g., snopes.com, politifact.com) have been developed. However, these websites require human expert analysis that takes a lot of time and effort.

Recent studies proposed systems to automatically detect false and inaccurate information. Rashkin et al. (2017) used linguistic information from the claims to differentiate between credible and non-credible claims. On the other hand, Popat et al. (2018) considered external evidence from articles that were retrieved from the web and which were relevant to the claims. Given that emotions play an important role for the propagation of rumors and false claims, Vosoughi et al. (2018) investigated true and false rumors on Twitter and found that false rumors triggered fear, disgust, and surprise in their replies, whereas the true rumors triggered joy, sadness, trust, and anticipation. However, they did not explore the effectiveness of emotions in automatic false information detection. Emotions were also considered by Ghanem et al. (2020) who explored their impact on the detection of the different types (i.e., propaganda, hoax, clickbait, and satire) of misinformation.

In this paper, we focus on understanding whether a claim made by a politician is credible or not. Different from previous works, we are interested to study the impact of emotional signals on differentiating between credible and non-credible claims. In particular, we focus on credibility assessment⁵ that aims to differentiate between credible and non-credible claims. Table 1 shows some examples of claims with their credibility label. From these examples, we observe that in case of non-credible claims there is an attempt to trigger emotions that are different compared to the credible claims. In particular, the non-credible claims contain words such as *killed* and *drowning* that can potentially trigger fear and anger. On the contrary, credible claims describe news using a neutral language.

In this study, we hypothesize that non-credible claims can cause different emotions and of different intensity to the readers compared to the credible claims. Therefore, emotional signals can play an important role in

TABLE 1 Example of labels with their credibility label

| Claim | Label |
|---|----------------|
| You cannot build a Christian church in Saudi Arabia | True |
| We see a quarter-billion dollars in a pension fund that needs to be funded at \$1.2 billion | Mostly true |
| We have killed lots of innocent people in the state of Texas | Mostly false |
| Far more children died last year drowning in their bathtubs than were killed accidentally by guns | Pants-on-fire! |

differentiating between them. If our hypothesis is correct, then a system that uses emotional signals for credibility detection could potentially be used by journalists as a first indicator to assess claims once they are made and hence minimize the negative consequences.

In this study, we explore different approaches for extracting emotional reactions from text. Then, we present *emoCred*, a long-short term memory (LSTM)-based approach that incorporates emotional signals and that can determine whether a claim is credible or not. In addition, we analyze which emotional signals and which terms are the most useful for the different classes of the credibility. This is an important information to better understand the task and if there are specific classes that are more challenging and they need more attention from credibility detection systems. Also, with this analysis it is possible to make observations regarding which terms are important per class that can further help in developing an effective credibility detection system. In particular, we focus on the following research questions:

- RQ1.** Can state-of-the-art deep learning approaches predict the ordinal level (i.e., low, average, high) of emotional reactions that are triggered by news posts?
- RQ2.** Can we use emotional signals from the claims to improve the performance of credibility assessment?
- RQ3.** Are the emotional signals from claims useful for a more fine-grained classification of credibility?

The contributions of this paper can be summarized as follows:

1. We explore different approaches for predicting the level of emotional reactions that a news post will trigger to users.
2. We present *emoCred*, a system that incorporates emotional signals from the text of the claim for credibility assessment.
3. We measure the effectiveness of *emoCred* in a fine-grained credibility assessment setting with six different credibility labels.

This paper extends and complements our previous works on emotion reaction prediction (Giachanou et al., 2018b) and credibility assessment (Giachanou, Gonzalo, & Crestani, 2019) by providing a far more complete presentation of the topic and a more thorough analysis of the results. It contains considerable extensions of the original papers that can be summarized as: (a) We compare the effectiveness of different neural networks on extracting and predicting the ordinal level of emotional reactions

that an article can trigger; (b) We provide a more extensive set of experiments on credibility detection; (c) We study the effect of the different hyper-parameters of emoCred on the credibility detection performance, and finally; (d) we provide a more thorough analysis on our results.

We evaluate emoCred on real-world datasets that contain political claims. We use claims published in Politifact⁶ and manually labeled by human experts. Our experiments are performed with two different settings: (a) binary classification (i.e., false, true), and (b) six-class classification (i.e., pants-on-fire, false, mostly false, half true, mostly true, true). Our results show that emoCred significantly outperforms the baselines in both settings, and therefore, emotional signals should be considered in the approaches that address the problem of credibility detection.

2 | RELATED WORK

In this section, we first present related work on fake news and credibility detection and then, we discuss prior work on emotion analysis.

2.1 | Fake news and credibility detection

Credibility detection can be viewed as a branch of the bigger research area of fake news detection that has recently attracted a lot of attention. Recently, a great number of the studies proposing approaches on fake news detection are based on deep learning methods (Barrón-Cedeno et al., 2018; Popat et al., 2018; Shu, Sliva, et al., 2017). One characteristic of deep learning approaches is their ability to capture information directly from the data. For example, Recurrent Neural Networks (RNN) are designed to recognize patterns in sequences of data such as text, speech, or numerical times series data. LSTM networks (Hochreiter & Schmidhuber, 1997) is a type of RNN network that has recently been applied for various problems such as fake news detection and sentiment analysis. Convolutional Neural Networks (CNNs) are another type of neural networks that use a variation of multilayer perceptrons designed to require minimal preprocessing (Krizhevsky et al., 2012). Although CNNs are very common for image classification, they have been also used a lot in text classification problems (Kim, 2014).

Shu, Sliva, et al. (2017) defines *fake news as news articles that are intentionally and verifiably false*. There are a lot of concepts related to fake news detection such as fact checking (Karadzhov et al., 2017; Thorne et al., 2018) and rumor detection (Buntain & Golbeck, 2017; Ma et al., 2016). Fact

checking mostly aims to detect true facts using different sources such as online articles or Wikipedia, whereas rumor detection aims to classify a post as rumor or non-rumor. A rumor is defined as *an item of circulating information whose veracity status is yet to be verified at the time of posting* (Zubiaga et al., 2018). Credibility assessment is another related task which aims to differentiate between credible and non-credible claims (Rashkin et al., 2017). Different evaluation campaigns have been organized recently to facilitate research on all these problems. FEVER (Thorne et al., 2018), CLEF Fact Checking (Nakov et al., 2018), Profiling Fake News Spreaders at PAN (Rangel et al., 2020) and SemEval rumor detection (Derczynski et al., 2017) are some examples of the various campaigns that have been organized over the last years. A comprehensive review on the problem of fake news detection is given by Shu, Sliva, et al. (2017) and Kumar and Shah (2018).

Existing approaches on fake news detection mainly rely on a wide range of information ranging from textual to visual and users' reactions to the source credibility. Textual information is represented using common linguistic features that include lexical features (e.g., terms, sequence of terms), and syntactic features (e.g., part-of-speech tags). In this direction, Zhou et al. (2020) have proposed a theory-driven model for fake news detection. Their method investigated news content at various levels: lexicon-level, syntax-level, semantic-level, and discourse-level. Another study from Pérez-Rosas et al. (2018) focused on the effectiveness of text-based features on false news detection. In their study, they explored the effectiveness of a wide range of text-based features such as punctuation, readability and syntax features on a dataset containing false information generated by Amazon Mechanical Turk workers and on another dataset from GossopiCop. Their experiments showed the effectiveness of those features and the challenge of achieving a good performance on a cross-domain setting. Another approach was presented by Shu, Sliva, et al. (2017) who proposed to exploit the relationships among publisher, news, and users engagements on social media to detect fake news, whereas Ruchansky et al. (2017) combined two different modules to detect fake news. The first module captured the temporal engagements of users using a RNN, whereas the second used a neural network and a user graph to assign a score to each user.

Regarding credibility detection, Rashkin et al. (2017) trained a LSTM model using linguistic features from the text of the claims to distinguish between credible and non-credible claims. They incorporated various linguistic features extracted using the Linguistic Inquiry and Word Count (LIWC) dictionary (Tausczik & Pennebaker, 2010) such as personal pronouns and swear words. In addition to text, Wang (2017) used different meta-data of the

claims such as the claim's author and subject whereas Potthast et al. (2018) proposed domain-specific linguistic features aligned to the news domain such as external links and number of paragraphs.

Apart from the text of the claim, other researchers considered external evidence from the web. For example, Popat et al. (2017) combined different information extracted from external resources such as stance detection and trustworthiness of the sources. A pipeline of supervised classifiers was used to combine the different sources of information. External evidence was also considered in another paper by Popat et al. (2018) who proposed an evidence-aware neural network model. To classify the claims as credible or not credible, they aggregated signals from the external evidence articles and the reliability of their sources. The evidence articles were relevant to the claims and were collected using a search engine. In addition, other researchers have explored the effectiveness of multimodal content on fake news detection. Giachanou et al. (2020) proposed a multimodal system that combined textual, visual, and semantic information, whereas Khattar et al. (2019) proposed the Multimodal Variational Autoencoder (MVAE) model based on bidirectional LSTMs and VGG-19 for the text and image representation to differentiate between fake and real news.

Apart from fake news and credibility detection, researchers have also tried to address a wide range of similar tasks. For example, Ghanem et al. (2020) explored the effectiveness of emotions in the detection of the different types of misinformation, whereas Tacchini et al. (2017) proposed a semi-supervised probabilistic approach to predict hoaxes using the Facebook like button. Castillo et al. (2011) presented different credibility features to determine the reliability of posts in Twitter, whereas Ma et al. (2016) proposed an RNN model for learning the hidden representations to capture the variation of contextual information of relevant posts over time with the aim to detect rumors in social media. Jin et al. (2016) presented a credibility propagation network for news verification based on stance. Finally, there have also been attempts to profile the users that tend to share fake news on social media (Giachanou et al., 2020; Shu, Wang, & Liu, 2017).

Different from previous studies, we are interested in investigating the effectiveness of emotional signals extracted from the text of the claim in credibility assessment. Our aim is to explore how much improvement we can achieve by using the emotional signals of the claims. In particular, we extract emotional signals from the claims and we then incorporate them into an LSTM neural network to explore their impact on credibility detection. Different from Popat et al. (2018) we do not use

evidence articles from the web. To the best of our knowledge, we are the first to consider the effectiveness of emotional signals for detecting credible and non-credible claims.

2.2 | Emotion analysis

Detecting the sentiment or the emotion that is expressed in a piece of text is a very challenging problem. Even if sentiment and emotion analysis are two different tasks, they are closely related and addressed with similar approaches. Sentiment analysis refers to classifying a document as expressing positive or negative sentiment toward a specific entity, whereas emotion analysis focuses on determining the emotions that are expressed in a document. To define the set of emotions, researchers have used theories from psychology. One of the most common is the Ekman's theory, according to which, there are six core emotions: fear, anger, disgust, sadness, happiness, and surprise (Ekman, 1992).

Early models for sentiment and emotion analysis include classification and lexicon-based approaches, whereas recently researchers have also considered deep learning approaches (Dos Santos & Gatti, 2014; Severyn & Moschitti, 2015). Classification based approaches rely on a number of different hand-crafted features to classify the documents as expressing a specific sentiment or emotion. Lexicon-based approaches rely on set of words whose presence usually imply a specific emotion (Taboada et al., 2011). Several sentiment and emotion lexicons have been developed to facilitate research in emotion analysis (Mohammad & Turney, 2013). Interested readers are referred to surveys that present a comprehensive review of those topics (Giachanou & Crestani, 2016; Zhang et al., 2018).

Apart from the emotions that are expressed in a document, the intensity level is also an important information. Mohammad and Bravo-Marquez organized a shared task on detecting the intensity of emotion felt by the author of a tweet (Mohammad & Bravo-Marquez, 2017). Some of the most common features that were used for the emotion intensity task were word embeddings and affective lexicons. One of the most common lexicons is the Affect Intensity Lexicon (Mohammad, 2018) that provides intensity scores and which is particularly helpful for the emotion intensity task. The best performance on the emotion intensity task was achieved by Goel et al. (2017) who proposed an ensemble of three different deep learning based approaches combined using a weighted average of the separate predictions. Giachanou et al. (2018a) predicted the ordinal level of emotional reactions (love, joy, surprise, sadness, anger) with

features that were extracted from the users' comments, such as when the first comment was published, to predict the emotional reactions of news posts.

Sentiment and emotions can play an important role in addressing different problems ranging from irony detection (Hernandez-Farías et al., 2016) to recommender systems (Zheng et al., 2016) and from stance detection (Mohammad et al., 2017) to online reputation polarity detection (Giachanou, Rosso, & Crestani, 2019; Giachanou et al., 2017). In addition, emotions can be crucial for fake news detection. Vosoughi et al. (2018) investigated a large corpus of tweets and found that false rumors trigger fear, disgust, and surprise in their replies, whereas the true rumors trigger joy, sadness, trust, and anticipation. Similar to previous works that explored the role of sentiment and emotions in addressing different problems and inspired by the work of Vosoughi et al. (2018), in this study we extract emotional signals from the claims to evaluate their effectiveness on the credibility detection problem.

3 | METHODOLOGY

In this section, we present our methodology for addressing the credibility detection problem. First, we present different approaches for extracting and estimating the emotional signals from the text of claims. Next, we present the *emoCred* neural network approach that incorporates emotional signals for detecting credible and non-credible claims.

3.1 | Emotional signals

A preliminary step of our methodology is to estimate the emotional signals from the claims. We consider three different approaches to calculate the emotional signals that are expressed in a claim: (a) a lexicon-based approach that is based on state-of-the-art emotional lexicons (*emoLexi*), (b) an approach that calculates the emotional intensity of the claims (*emoInt*), and (c) an approach that assigns an intensity level to the claims and which represents the number of emotional reactions that the claim triggered to the readers (*emoReact*).

3.1.1 | emoLexi approach

The first approach is straightforward and is based on emotional lexicons. According to this approach, if specific words appear in a sentence, then the sentence expresses a specific emotion. For example, a sentence that contains the word *afraid* expresses *fear* whereas the word *excited*

conveys *joy* and *excitement*. More formally, let us consider a list of emotional words $\mathcal{E} = [t_{e_1}, t_{e_2}, \dots, t_{e_E}]$ that convey a specific emotion e . Then the *emoLexi* approach considers the number of emotional words that appear in the text of the claim. This process can be described as:

$$s(c, e) = \sum_{t \in c} f(t, c) \forall t \in \mathcal{E}$$

where $f(t, c)$ is the frequency of the term t in the claim c . After having calculated the emotional scores for different emotions, we can create the *esv* vector for the claim c as:

$$esv = \{s(c, e_1), s(c, e_2), \dots, s(c, e_E)\}$$

where $E = \{e_1, e_2, \dots, e_E\}$ is a list of E emotions.

3.1.2 | emoInt approach

For the *emoInt* approach, we consider an emotional intensity lexicon that contains a list of words that convey a specific emotion together with a score that represents the emotional intensity of the specific word. The emotional intensity of a claim c regarding a specific emotion e is calculated as:

$$s(c, e) = \sum_{t \in c} s_{int}(t, e) \forall t \in \mathcal{E}$$

where t is a term in a claim c and $s_{int}(t, e)$ is the intensity score of the term t toward the emotion e given the emotional intensity lexicon.

3.1.3 | emoReact approach

The *emoReact* approach aims to predict the ordinal level of emotional reactions that will be triggered to the readers (Giachanou et al., 2018b). This approach predicts for each claim the probability of leading to any of the three intensity levels (low, average, high) for each of the five reactions love, joy, surprise, sadness, and anger. In (Giachanou et al., 2018b) we proposed different textual features to predict the ordinal level of reactions using an SVM-classifier. In this paper, we additionally explore the use of different neural networks to understand which one performs better for the prediction of the reactions' ordinal level. In particular, we explore the following approaches: (a) a standard SVM bag of words approach, (b) an LSTM neural network, (c) a bi-LSTM neural network, and (d) a CNN network. All the neural networks are trained with pretrained word embeddings.

3.2 | emoCred credibility detection

We present EmoCred, an approach based on a LSTM neural network that incorporates emotional signals to differentiate between credible and non-credible claims. EmoCred takes as an input word embeddings from the text of the claims and a vector of emotional signals. Figure 1 gives an overview of the architecture of our model. To introduce our methodology more formally, let us consider a collection of N claims noted as C_N . Then, the representation of a claim c of length l is $[w_1, w_2, \dots, w_l]$ where $w_l \in \mathbb{R}^d$ and \mathbb{R}^d is the d -dimensional word embedding of the l th word in the input claim c .

As a second input, we consider emotional signals that are extracted from the text of the claim. In the previous sections, we described our methodology for extracting the emotional signals from the text of the claims. Given a vector of emotional signals esv , we first use a fully connected layer to output a proper representation for our neural network noted as d_e :

$$d_e = \text{relu}(W_e(esv) + b_e)$$

where W_e and b_e are the weight matrix and bias, respectively.

We then combine the claim word embeddings wc and the output d_e of the dense layer which is a representation of the emotional signals. To achieve this, we concatenate them with one fully connected layer as follows:

$$d_m = \text{relu}(W_m(wc \oplus d_e) + b_m)$$

where W_m and b_m are the weight matrix and bias, respectively.

To generate the final credibility label for a given claim regarding the binary classification, we apply the *sigmoid* function to the final representation as follows:

$$S_{sig} = \text{sigmoid}(d_m)$$

Regarding the six-class classification, we apply the *softmax* function to the final representation as follows:

$$S_{soft} = \text{softmax}(d_m)$$

4 | EXPERIMENTS

In this section, we describe the datasets and the experimental process we followed to conduct our experiments.

4.1 | Datasets

For the ordinal level classification of the emotional reactions, we use data from Facebook. The dataset is the same we used in Giachanou et al. (2018b) for determining the emotional triggers of news posts. The dataset contains 26,560 news posts that span from April 2016 to September 2017 crawled from New York Time Facebook page together with the actual number of emotional reactions that they triggered.

To evaluate our approaches on credibility detection we use data from PolitiFact that is a fact checking website where the credibility of different claims is investigated. We use two different Politifact datasets presented in two different studies (Popat et al., 2018; Rashkin et al., 2017). We decided not to merge the two datasets so we can use the predefined train, test, and development sets. The datasets do not overlap.

The first dataset⁷ (Politifact-1) was used in the study presented by Popat et al. (2018). The dataset contains 3,568 claims of which 1,867 are credible and 1,701 are not credible claims. The dataset also contains snippets

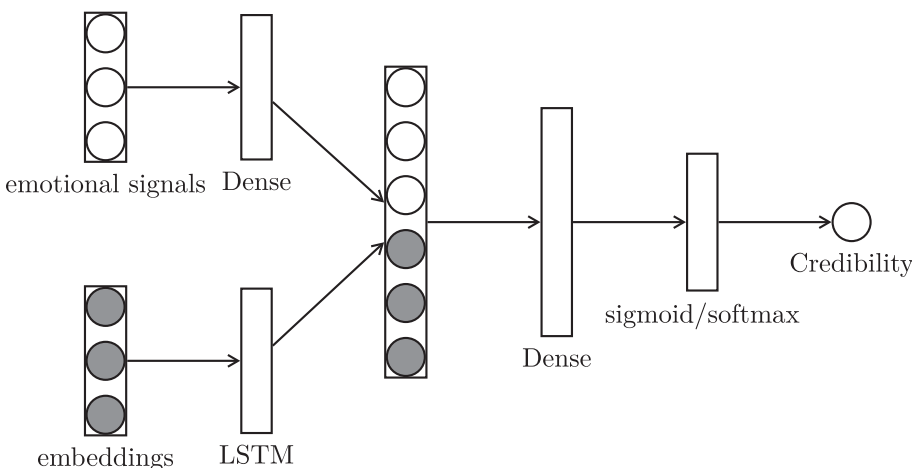


FIGURE 1 EmoCred neural network architecture for credibility assessment

from 29,556 articles that correspond to a total of 336 article sources.

The second dataset (Politifact-2) was presented by Rashkin et al. (2017). This dataset consists of 2,575 training, 712 development and 1,074 test statements. Both datasets contain the text of the claims, the speaker, and the credibility rating of each claim. There are six different credibility ratings: {true, mostly true, half true, mostly false, false and pants-on-fire}. We treat the problem in two different settings: (a) as a binary classification, and (b) as a six-class classification. For the binary classification setting, we follow the same process followed by previous studies (Popat et al., 2018; Rashkin et al., 2017). In particular, we combine *true*, *mostly true*, and *half true* labels into one class label (i.e., true) and the rest (i.e., *mostly false*, *false*, and *pants-on-fire*) into false.

Table 2 shows the distribution of the labels on the datasets. We observe that Politifact-1 is balanced regarding the binary classification. However, in the six-class classification the number of labels are not balanced. In particular, the Pants-on-Fire class contains the less data. The data in Politifact-2 are imbalanced for both the two-class and the six-class classification. Similar to Politifact-1, the class with the less data is the Pants-on-Fire class. The fact that the data are not balanced may create learning challenges for our methodologies.

4.2 | Emotional lexicons

In order to calculate the emotional signals in the claims with the *emoLexi* approach, we use the following emotional lexicons:

- EmoLex (Mohammad & Turney, 2013): This lexicon contains the associations of 14,181 words with eight emotions (i.e., anger, anticipation, disgust, fear, joy, sadness, surprise, trust).
- SentiSense (de Carrillo Albornoz et al., 2012): A concept based affective lexicon that contains associations between WordNet concepts and emotional categories.
- EmoSenticNet (Poria et al., 2013): A lexical resource that contains emotional labels for SenticNet concepts.

We use the NRC Affect Intensity Lexicon (Mohammad, 2018) for the *emoInt* approach. The lexicon contains around 6,000 entries associated with a real number that represents the intensity of the term with regards to the emotion. We calculate the intensities in respect to four basic emotions: anger, fear, joy, and sadness.

The *emoReact* approach detects the emotional reactions signals using a real corpus crawled by Facebook which was described in the previous section. We compare different neural networks for the prediction of emotional reactions. In particular, we train a LSTM, a bi-LSTM, and a CNN network. Among these models, we use the CNN neural network trained on the Facebook posts to predict the intensity of the emotional reactions in the claims since it achieved the best result among the neural networks. More specifically, for each of the claim, we predict the probability to lead to any of the three different intensities (low, average, high) of five different emotional reactions (love, joy, sadness, surprise, anger). We use the pretrained GloVe Wikipedia 6B word embeddings (Pennington et al., 2014) to initialize the word embeddings.

4.3 | Experimental setup

We split Politifact-1 as following: 10% of the data is kept for validation to tune the parameters of the models, 20% is kept for test and the remaining 70% to train the models. For our experiments on Politifact-2 we use the training, test and development sets already provided in the corpus. We use the pretrained GloVe Wikipedia 6B word embeddings (Pennington et al., 2014) to initialize our word embeddings.

Regarding the emotional predictions problem, we use scikit-learn for the SVM and Keras with a Tensorflow backend for the LSTM, bi-LSTM, and CNN networks. In particular, we use a linear kernel for the SVM and the One-vs-the-rest (OvR) multilabel strategy. Table 3 shows the hyper-parameters we used for the different networks on the emotional prediction.

We use Keras with a Tensorflow backend to implement our system. All the neural networks models are trained with a learning rate of 0.002 to minimize binary

TABLE 2 Label distribution of the datasets

| Dataset | Classes | True | Mostly true | Half true | Mostly false | False | Pants-on-fire |
|--------------|-----------|--------|-------------|-----------|--------------|--------|---------------|
| Politifact-1 | Two-class | 52.05% | – | – | – | 47.95% | – |
| | Six-class | 14.46% | 17.69% | 19.90% | 16.64% | 20.75% | 10.56% |
| Politifact-2 | Two-class | 60.40 | – | – | – | 39.60 | – |
| | Six-class | 19.37% | 20.46% | 20.57% | 14.13% | 17.97% | 7.49% |

| | Lstm units | Activation | Optimizer | Epochs | Batch size |
|---------|-------------------|------------|-----------|--------|------------|
| LSTM | 64 | Relu | Rmsprop | 32 | 64 |
| Bi-LSTM | 64 | Relu | Rmsprop | 32 | 64 |
| CNN | 3 (filters sizes) | Relu | Adam | 10 | 64 |

TABLE 3 Parameter optimization for the emotional prediction for the different tested systems

| Dataset | Method | Units | Epochs | Dropout | Batch size |
|--------------|---------------------------|-------|--------|---------|------------|
| Politifact-1 | LSTM-text, LSTM-text-LIWC | 50 | 48 | 0.5 | 128 |
| | emoCred-emoLexi | 128 | 16 | 0.5 | 64 |
| | emoCred-emoInt | 128 | 48 | 0.5 | 64 |
| | emoCred-emoReact | 64 | 32 | 0.3 | 32 |
| Politifact-2 | LSTM-text, LSTM-text-LIWC | 50 | 48 | 0.5 | 64 |
| | emoCred-emoLexi | 128 | 16 | 0.5 | 32 |
| | emoCred-emoInt | 64 | 32 | 0.5 | 64 |
| | emoCred-emoReact | 128 | 72 | 0.3 | 64 |

TABLE 4 Hyper-parameters on the binary classification

| Dataset | Method | Units | Epochs | Dropout | Batch size |
|--------------|---------------------------|-------|--------|---------|------------|
| Politifact-1 | LSTM-text, LSTM-text-LIWC | 50 | 12 | 0.2 | 128 |
| | emoCred-emoLexi | 100 | 32 | 0.3 | 128 |
| | emoCred-emoInt | 75 | 12 | 0.3 | 64 |
| | emoCred-emoReact | 175 | 24 | 0.5 | 64 |
| Politifact-2 | LSTM-text, LSTM-text-LIWC | 150 | 12 | 0.2 | 64 |
| | emoCred-emoLexi | 150 | 32 | 0.3 | 128 |
| | emoCred-emoInt | 100 | 24 | 0.3 | 128 |
| | emoCred-emoReact | 150 | 24 | 0.5 | 128 |

TABLE 5 Hyper-parameters on the six-class classification

cross-entropy loss. We use L2 regularizers with the fully connected layer. We used the hyperopt library⁸ to search for the best parameters on the development sets. Tables 4 and 5 show the hyper-parameters we used for the different settings and approaches on the binary and six-class classification respectively.

We report accuracy and macro F1-metric for the evaluation of the models. We compare the performance of EmoCred with that of LSTM-text and LSTM-text-LIWC approach proposed by Rashkin et al. (2017) and standard SVM classifier trained on bag of words. This approach is based on a LSTM network trained using only word embeddings from the text of the claims, initialized using the pre-trained word embeddings. Finally, we use the McNemar test to measure the statistical difference (McNemar, 1947).

5 | RESULTS

In this section we present the results of our study. First, we present the results on the ordinal classification of the

emotional reactions. Next, we present the results on the credibility detection problem.

5.1 | Emotional reactions intensity prediction

Table 6 summarizes the results of our experiments regarding the emotional reactions prediction problem. From the results we observe that SVM performs better than deep learning approaches. In particular, SVM outperforms all the deep learning approaches by a large margin. Among the neural networks approaches we observe that the CNN-based model manages to obtain the best performance with regards to F1-metric. On the contrary, LSTM achieved the lowest performance.

The results on the emotional reaction problem suggest that for this specific problem standard supervised approaches are likely to outperform state-of-the-art neural networks. However, more experiments are needed to reach more robust conclusions. These are beyond the aim

TABLE 6 Performance results on the prediction of the ordinal level of the emotional reactions

| Method | Recall | Precision | F1-score |
|---------|--------|-----------|----------|
| SVM | 0.429 | 0.527 | 0.472 |
| LSTM | 0.271 | 0.393 | 0.307 |
| Bi-LSTM | 0.296 | 0.415 | 0.322 |
| CNN | 0.365 | 0.425 | 0.377 |

TABLE 7 Performance results of EmoCred approach when using different approaches for generating the emotional signals on the binary classification

| Dataset | Method | Accuracy | F1-score |
|--------------|------------------|----------|----------|
| Politifact-1 | SVM-text | 0.551 | 0.551 |
| | LSTM-text | 0.551 | 0.549 |
| | LSTM-text-LIWC | 0.541 | 0.532 |
| | EmoCred-emoLexi | 0.608 | 0.602* † |
| | EmoCred-emoInt | 0.604 | 0.602* † |
| | EmoCred-emoReact | 0.613 | 0.616* † |
| Politifact-2 | SVM-text | 0.568 | 0.561 |
| | LSTM-text | 0.597 | 0.567 |
| | LSTM-text-LIWC | 0.582 | 0.546 |
| | EmoCred-emoLexi | 0.621 | 0.606* † |
| | EmoCred-emoInt | 0.628 | 0.586† |
| | EmoCred-emoReact | 0.617 | 0.603* † |

Notes: The symbols * and † indicate statistically significant improvement over the SVM-text and LSTM-text approaches, respectively.

of this study that is to explore the effectiveness of emotional signals on credibility detection. For the rest of the experiments, we use CNN network to determine the emotional reactions with the *emoReact* approach. The decision is based on the fact that the model is trained on Facebook articles and then applied on political claims. Therefore, we believe that a neural network is more appropriate for such cross-domain setting.

5.2 | Credibility detection prediction

Table 7 summarizes the results of our experiments on the binary classification of credibility. From the results, we observe that EmoCred outperforms both the SVM and the LSTM baselines by a large margin. In most of the cases, we observe that incorporating emotional signals into LSTM significantly improves the performance for credibility detection compared to the baselines. Regarding the Politifact-1 dataset, the best performance is achieved with the *emoReact* approach.

TABLE 8 Performance results of EmoCred approach when using different approaches for generating the emotional signals on the six-class classification

| Dataset | Method | Accuracy | F1-score |
|--------------|------------------|----------|----------|
| Politifact-1 | SVM-text | 0.215 | 0.198 |
| | LSTM-text | 0.204 | 0.201 |
| | LSTM-text-LIWC | 0.203 | 0.198 |
| | EmoCred-emoLexi | 0.219 | 0.216* † |
| | EmoCred-emoInt | 0.237 | 0.224* † |
| | EmoCred-emoReact | 0.225 | 0.219* † |
| Politifact-2 | SVM-text | 0.182 | 0.185 |
| | LSTM-text | 0.202 | 0.196 |
| | LSTM-text-LIWC | 0.192 | 0.191 |
| | EmoCred-emoLexi | 0.233 | 0.211* † |
| | EmoCred-emoInt | 0.219 | 0.214* † |
| | EmoCred-emoReact | 0.223 | 0.212* † |

Notes: The symbols * and † indicate statistically significant improvement over the SVM-text and LSTM-text approaches, respectively.

From the table we observe that EmoCred-emoReact manages to significantly outperform LSTM-text by 12.20% in terms of the macro F1-score. This is a very interesting result because *emoReact* is trained on different data which were crawled from Facebook. In fact, even though the model was trained on data from a different domain, it seems that still the emotional features are very helpful for credibility detection. In case of *emoInt* and *emoLexi*, we observe that the two approaches obtain a similar performance. Both EmoCred-emoLexi and EmoCred-emoInt manage to significantly outperform LSTM-text by 9.65% and SVM-text by 9.25%. Similarly, EmoCred manages to outperform LSTM-text baseline on Politifact-2. In this case, the best performance is obtained with the *emoLexi* approach that significantly outperforms the LSTM-text baseline by 6.5%. In addition, we observe that the LSTM-text-LIWC obtains the lowest performance compared to SVM-text and LSTM-text that is compatible with the results presented by Rashkin et al. (2017).

Table 8 summarizes the results of our experiments on the six-class classification. From the results, we observe that emoCred significantly outperforms LSTM-text and SVM-text baselines by a large margin. The best performance on Politifact-1 is achieved with the *emoCred-emoInt* that manages to outperform LSTM-text by 11.44%. Similarly on Politifact-2, *emoCred-emoInt* significantly outperforms the LSTM-baseline by 9.18%. The results show that the emotional signals are very important also with a more fine-grained classification on credibility that is more challenging than the binary classification.

From the results, we also observe that the SVM-text achieves a lower performance compared to the LSTM-text and to emoCred. This result can be justified by the fact that deep learning approaches perform better than statistical learning-based approaches in many classification tasks since they can learn high-level features from data and they are not trained only on hand-crafted features.

In addition, we conducted a further analysis on the emotional signals to understand which emotions contributed more to the performance. With this purpose, we run the EmoCred-emoLexi on the PolitiFact-1 dataset using groups of emotions instead of all the emotions. More specifically, we choose to run it once using only fear, disgust, and surprise (group1) and once using only joy, sadness, trust, and anticipation (group2). This decision is inspired from the study by Vosoughi et al. (2018) who found that false rumors cause fear, disgust, and surprise whereas true rumors trigger joy, sadness, trust, and anticipation.

Table 9 shows the results of the EmoCred-emoLexi when we used two different groups of emotions instead of all the emotions. From the table, we observe that EmoCred obtained better results with emotions that are found in text with true information (i.e., joy, sadness, trust, and anticipation) compared to emotions found in text with false information (i.e., fear, disgust, and surprise). Finally, EmoCred-emoLexi (group2) significantly

TABLE 9 Performance results of EmoCred-emoLexi approach when using different groups of emotional signals

| Method | Accuracy | F1-score |
|--------------------------|----------|----------|
| LSTM-text | 0.551 | 0.549 |
| EmoCred-emoLexi (group1) | 0.589 | 0.587 |
| EmoCred-emoLexi (group2) | 0.605 | 0.599* |

Notes: Group1 includes fear, disgust, and surprise, whereas group2 joy, sadness, trust, and anticipation. A star (*) indicates statistically significant improvement over the LSTM-text approach.

improves the baseline in contrast to EmoCred-emoLexi (group1) that also improves the baseline but not significantly.

6 | ANALYSIS

In this section, we perform further analysis on our results with the purpose of better understanding the findings of the study. First, we analyze the emotional reactions prediction results. Then, we analyze our results regarding the credibility detection. Finally, we show the effect of the LSTM units parameter on the emoCred performance.

6.1 | Emotional reaction prediction

Figure 2 shows the recall, precision and F1 scores of the SVM classifier on the predicting the ordinal level of emotional reactions for each emotion reaction and for each class. From the results we observe that the lowest performance is achieved on the average class for all the emotion reactions. The low and high classes outperform the average class by a large margin. Regarding the performance across the emotional reactions, love and surprise achieve a lower performance than joy, sadness, and anger.

Figure 3 shows the recall, precision and F1 scores of the CNN neural network on the emotion prediction for each emotion reaction and for each class. In the case of CNN we observe that the results do not follow an obvious pattern across the different classes. Regarding the F1-scores we observe that the three classes perform similar or have small differences. The highest performance across the emotional reactions and the ordinal classes is achieved for the high class of the joy emotion. Regarding precision scores, we observe that the average class

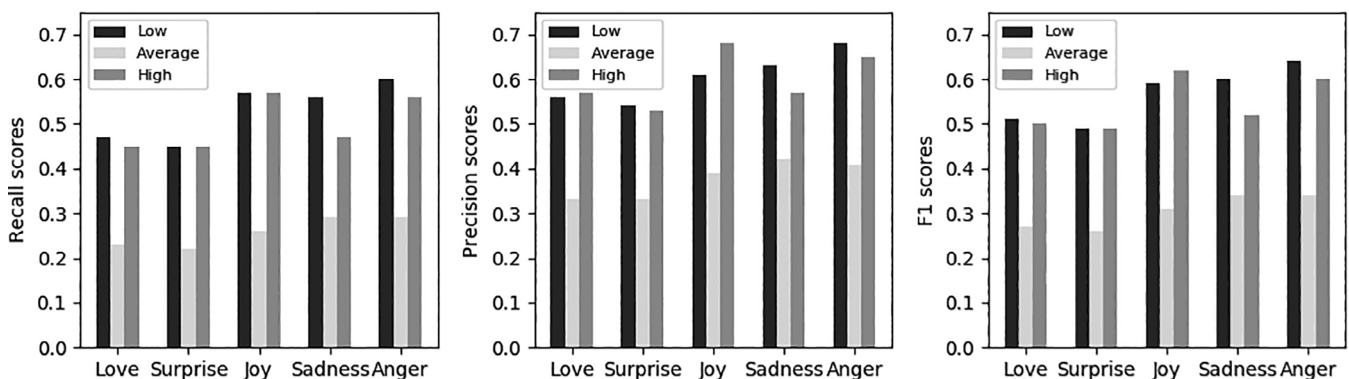


FIGURE 2 Recall, precision, and F1 scores of the SVM classifier per each class

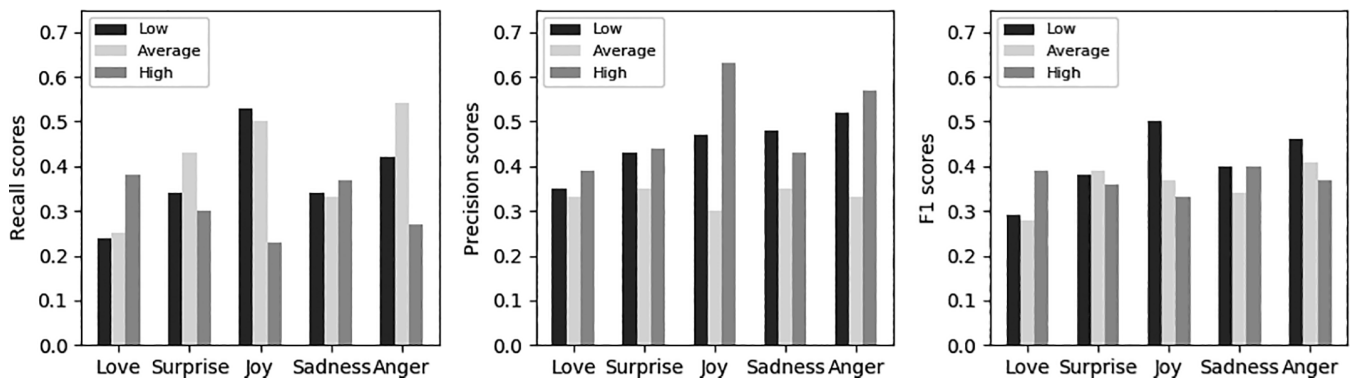


FIGURE 3 Recall, precision, and F1 scores of the CNN network per each class

achieves the lowest performance compared to the low and high classes across all the five emotional reactions. In general, we can say that even though SVM outperforms CNN, the predictions across the classes are more equally distributed when a CNN neural network is used. The benefits for CNN to be more equally distributed for the three classes is that in this way the risk of learning to predict only on few information that is only useful for the discrimination of the two classes and which can potentially lead to overfitting is minimized.

6.2 | Credibility detection analysis

In this section, we perform different analysis on the credibility detection problem. First, we analyze the effect of terms on the performance. Next, we explore the effect of LSTM units on emoCred performance.

6.2.1 | Effect of terms on performance

To understand which words are the most important for credibility detection and if there are any emotional terms in these words, we calculate the coefficient of correlation when the SVM-text baseline is used. In particular, Figure 4 shows the correlation coefficient of the terms per each of the six classes when the SVM-text approach is used on the Politifact-1 dataset. From the figures, we observe that there are some terms that express emotions such as the terms *crime* in the *Pants-on-Fire* class and the terms *worse* and *accused* in the *False* class. In addition, we observe that the highest correlation coefficients are about the class *False*.

Also, we perform an analysis on the emoCred results to explore the performance across the six different credibility classes. Figure 5 shows the performance of emoCred per class on credibility detection when *emoLexi*,

emoInt and *emoReact* approaches are used for calculating the emotional signals. In general, we observe that the best performance is achieved for the *False* and *Mostly True* classes. A very interesting observation is that the lowest performance when the emoCred is used is achieved for the *Pants-on-Fire* class. One reason for that may be the fact that these claims are intentionally written to confuse the readers and contain a mixture of false and true information. The low performance on the *Pants-on-Fire* also emphasizes on the difficulty of detecting non-credible claims since they are written in a way to be similar to the true claims.

6.2.2 | Effect of emoCred parameters on performance

Finally, we analyze the impact of LSTM units on the credibility detection performance. In particular, Figure 6 shows the effect of the LSTM units on the credibility assessment performance on Politifact-1 per class regarding the six-class classification. From the figures, we observe that in most of the cases lowest performances are obtained for the classes *Pants-on-Fire* and *True*. Also, we observe that there are not big differences across the different LSTM units. However, the optimal number of LSTM-units is different across the classes and the approaches. This raises the problem of setting up LSTM for approaches that are only partially known.

7 | CONCLUSIONS AND FUTURE WORK

In this paper, we presented an effective methodology to detect credible and non-credible claims. Starting from the assumption that false information cause different emotions with different intensity on the readers compared to

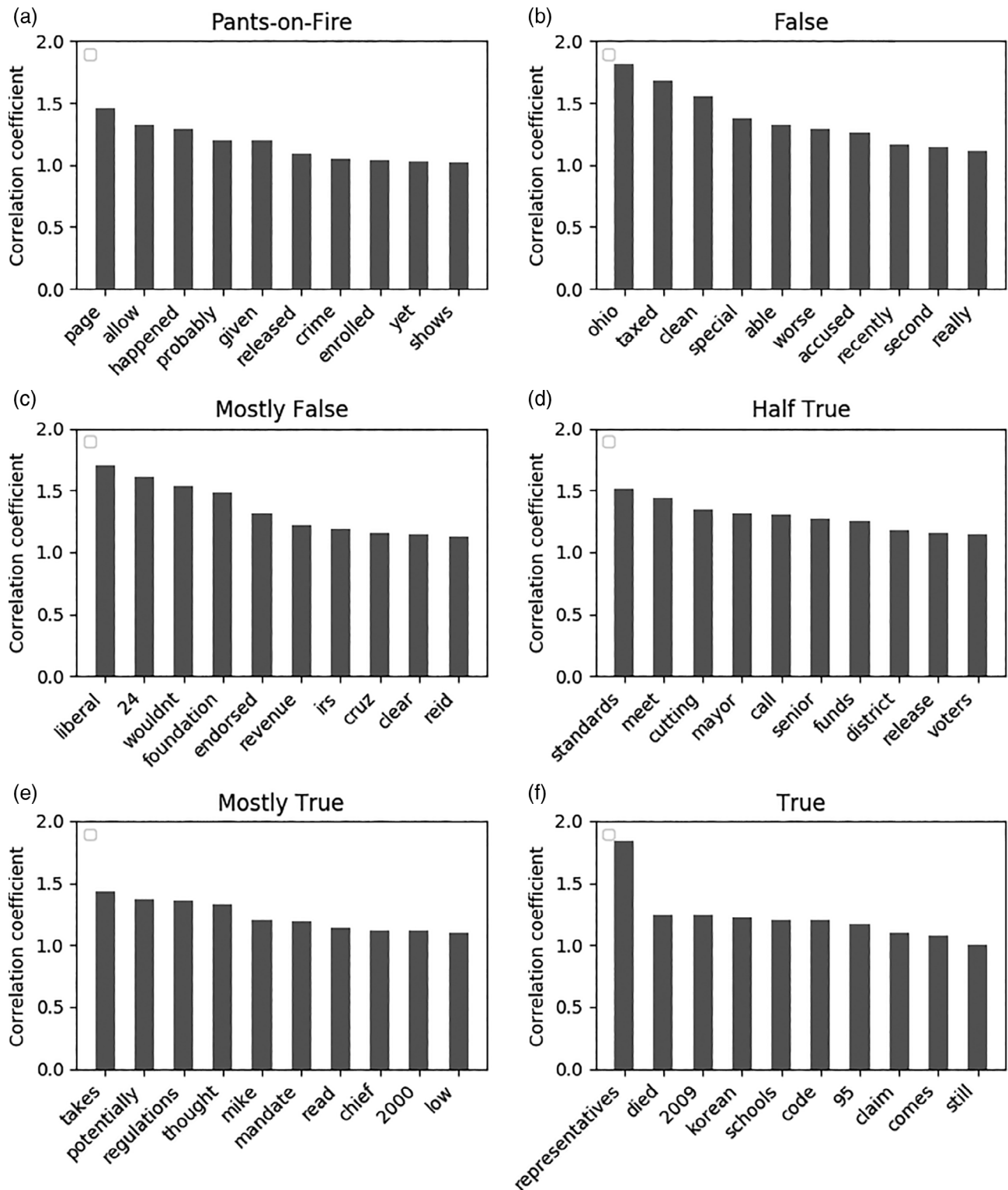


FIGURE 4 Correlation coefficients of terms per each class regarding the Politifact-1 dataset

true information, we explored the effectiveness of emotional signals on credibility detection. First, we explored the effectiveness of deep learning for the emotion reaction intensity. Then, we proposed emoCred, an approach based on a LSTM network that incorporates emotional

signals to detect credible claims. We explored different methodologies for calculating the emotional signals of the claims.

Our experiments on the emotional reactions prediction showed that a standard supervised approach

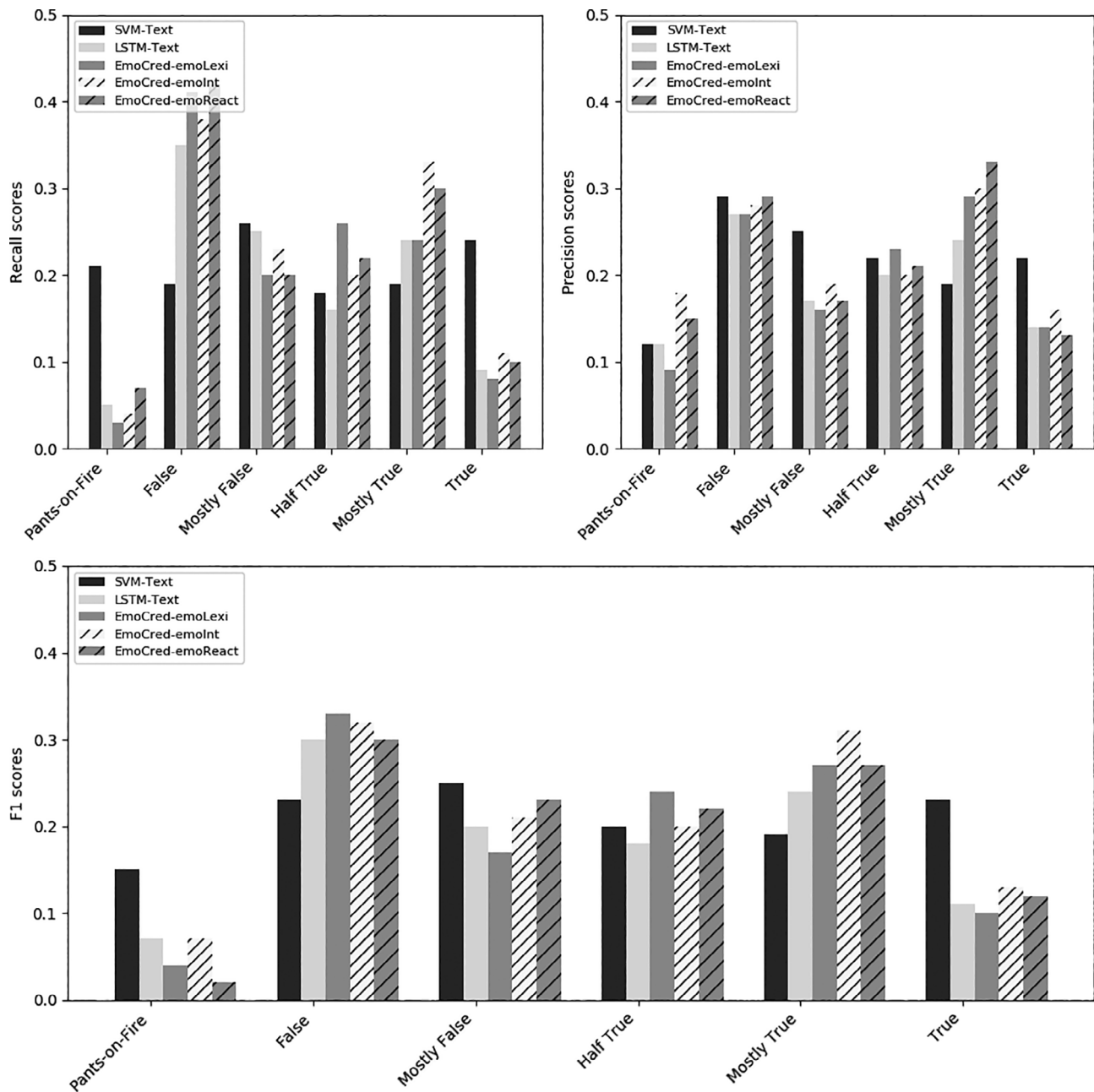


FIGURE 5 Performance scores of baselines and emoCred approaches per class on Politifact-1

outperforms the deep learning approaches. Also, the results showed that the most difficult class to predict is the one that refers to the average intensity level. Among the deep learning approaches, CNN managed to perform better compared to bi-LSTM and LSTM networks. Most importantly, the results of our experiments on credibility detection strongly support our initial hypothesis that emotional signals can be effectively incorporated into a LSTM-neural network and help in classifying credible and non-credible posts. The presented model, emoCred, managed to significantly improve the performance of the baseline approaches on credibility detection by a large

margin. Also, we showed that the most challenging classes were the *Pants-on-Fire* and the *True* ones.

Even if our study can provide valuable insights regarding the emotions that are expressed in credible and non-credible claims and their automated detection, there are some limitations. One limitation of the current work is the lexicon-based approach for detecting the emotions in the text. Although lexicon-based approaches based on general lexicons have the advantage of being domain independent, they are still language-dependent. Similar to all the automated tools that are used for predictions, the lexicon-based

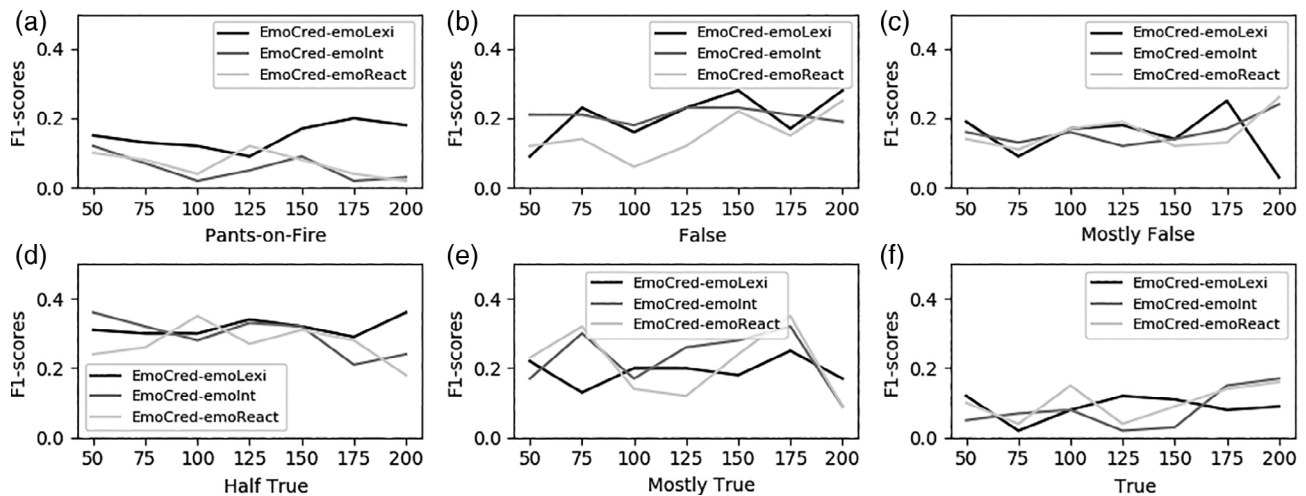


FIGURE 6 Effect of LSTM units on the emoCred credibility assessment performance on Politifact-1

approach is also subject to errors. That means that some of the predictions regarding the emotions expressed in the text are not correct. However, given the amount of data and the subjectivity of the task, manual annotation was not possible in our case.

In the future, we would like to explore the reasons why emoCred achieves a lower performance score on Pants on Fire and True classes compared to the rest of the classes by exploring the existence of any specific patterns in the misclassifications across the different classes. In addition, we plan to explore the effectiveness of ensemble neural networks on credibility detection and the role of the evidence articles (e.g., articles relevant to the claim retrieved from the web) and their emotional signals on credibility detection. In particular, we plan to extract emotions expressed in the evidence documents and explore their effectiveness on fake news and credibility detection. Finally, we plan to focus on intention detection that aims to understand whether a piece of fake news was shared with the intention to harm or not. We plan to modify emoCred and explore the effectiveness of extracting emotions and other textual features on the intention detection problem.

ACKNOWLEDGMENTS

Anastasia Giachanou was supported by the SNSF Early Postdoc Mobility grant P2TIP2_181441 under the project *Early Fake News Detection on Social Media*, Switzerland. The work of Paolo Rosso was partially funded by the Spanish MICINN under the research project MISIMIS-FAKENHATE on MISinformation and MIScommunication in social media: FAKE news and HATE speech (PGC2018-096212-B-C31) and by the Generalitat

Valenciana under the research project DeepPattern (PROMETEO/2019/121).

ORCID

Anastasia Giachanou  <https://orcid.org/0000-0002-7601-8667>

Fabio Crestani  <https://orcid.org/0000-0001-8672-0700>

ENDNOTES

- ¹ In this paper, we use the term fake news to refer to all the different types of false and inaccurate information including among others rumors, false claims, conspiracy theories, and clickbaits.
- ² <http://nymag.com/selectall/2016/11/donald-trump-won-because-of-facebook.html>
- ³ <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>
- ⁴ <https://www.weforum.org/agenda/2017/08/scientists-can-vaccinate-against-the-post-truth-era>
- ⁵ In this paper, we refer to the *credibility assessment problem* also with the phrase *credibility detection* interchangeably.
- ⁶ www.politifact.com
- ⁷ <https://www.mpi-inf.mpg.de/dl-cred-analysis/>
- ⁸ <https://github.com/hyperopt/hyperopt>

REFERENCES

- Barrón-Cedeno, A., Elsayed, T., Suwaileh, R., Márquez, L., Atanasova, P., Zaghouni, W., Kyuchukov, S., Da San Martino, G., and Nakov, P. (2018). Overview of the CLEF-2018 checkthat! lab on automatic identification and verification of political claims. task 2: Factuality. In *CLEF 2018 Working Notes. Working Notes of CLEF 2018—Conference and Labs of the Evaluation Forum*.

- Buntain, C., & Golbeck, J. (2017). Automatically identifying fake news in popular twitter threads. In *Proceedings of the 2017 IEEE International Conference on Smart Cloud* (pp. 208–215). IEEE.
- Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web* (pp. 675–684). ACM.
- de Carrillo Albornoz, J., Plaza, L., & Gervás, P. (2012). SentiSense: An easily scalable concept-based affective lexicon for sentiment analysis. In *Proceedings of the 8th International Conference on Language Resources and Evaluation* (pp. 3562–3567). ELRA.
- Derczynski, L., Bontcheva, K., Liakata, M., Procter, R., Hoi, G. W. S., & Zubiaga, A. (2017). SemEval-2017 task 8: Rumoureval: Determining rumour veracity and support for rumours. In *Proceedings of the 11th International Workshop on Semantic Evaluation* (pp. 69–76).
- Dos Santos, C., & Gatti, M. (2014). Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 69–78). ACL.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3/4), 169–200.
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104.
- Ghanem, B., Rosso, P., & Rangel, F. (2020). An emotional analysis of false information in social media and news articles. *ACM Transactions on Internet Technology*, 20(2), 1–18.
- Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys*, 49(2), 28:1–28:41.
- Giachanou, A., Gonzalo, J., & Crestani, F. (2019). Propagating sentiment signals for estimating reputation polarity. *Information Processing and Management*, 56(6), 102079.
- Giachanou, A., Gonzalo, J., Mele, I., & Crestani, F. (2017). Sentiment propagation for predicting reputation polarity. In *Proceedings of the 39th European Conference on Advances in Information Retrieval* (pp. 226–238). Springer.
- Giachanou, A., Rosso, P., & Crestani, F. (2019). Leveraging emotional signals for credibility detection. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM.
- Giachanou, A., Rosso, P., Mele, I., & Crestani, F. (2018a). Early commenting features for emotional reactions prediction. In *Proceedings of the 25th International Symposium on String Processing and Information Retrieval* (pp. 168–182).
- Giachanou, A., Rosso, P., Mele, I., & Crestani, F. (2018b). Emotional influence prediction of news posts. In *Proceedings of the 12th International AAAI Conference on Web and Social Media* (pp. 592–595). AAAI Press.
- Giachanou, A., Zhang, G., & Rosso, P. (2020). Multimodal fake news detection with textual, visual and semantic information. In *Text, speech, and dialogue* (pp. 30–38). Springer.
- Goel, P., Kulshreshtha, D., Jain, P., & Shukla, K. K. (2017). Prayas at emoint 2017: an ensemble of deep neural architectures for emotion intensity prediction in tweets. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 58–65). ACL.
- Hernandez-Farías, D. I., Patti, V., & Rosso, P. (2016). Irony detection in twitter: The role of affective content. *ACM Transactions on Internet Technology*, 16(3), 19:1–19:24.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Jin, Z., Cao, J., Zhang, Y., & Luo, J. (2016). News verification by exploiting conflicting social viewpoints in microblogs. In *Proceedings of the 13th AAAI Conference on Artificial Intelligence*. AAAI Press.
- Karadzhev, G., Nakov, P., Márquez, L., Barrón-Cedeño, A., & Koychev, I. (2017). Fully automated fact checking using external sources. In *Proceedings of the International Conference Recent Advances in Natural Language Processing* (pp. 344–353).
- Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). Mvae: Multimodal variational autoencoder for fake news detection. In *The World Wide Web Conference* (pp. 2915–2921). Association for Computing Machinery.
- Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1746–1751). ACL.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097–1105). ACM.
- Kumar, S., & Shah, N. (2018). False information on web and social media: A survey. *arXiv preprint arXiv:1804.08559*.
- Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K.-F., & Cha, M. (2016). Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence* (pp. 3818–3824). AAAI Press.
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2), 153–157.
- Mohammad, S. M. (2018). Word affect intensities. In *Proceedings of the 11th International Conference on Language Resources and Evaluation* (pp. 174–183). ELRA.
- Mohammad, S. M., & Bravo-Marquez, F. (2017). WASSA-2017 shared task on emotion intensity. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (p. 34–49).
- Mohammad, S. M., Sobhani, P., & Kiritchenko, S. (2017). Stance and sentiment in tweets. *Transactions on Internet Technology*, 17(3), 26:1–26:23.
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3), 436–465.
- Nakov, P., Barrón-Cedeño, A., Elsayed, T., Suwaileh, R., Márquez, L., Zaghoulani, W., Atanasova, P., Kyuchukov, S., Da San Martino, G. (2018). Overview of the CLEF-2018 CheckThat! lab on automatic identification and verification of political claims. In *Proceedings of the 9th International Conference of the CLEF Association: Experimental IR Meets Multilinguality, Multimodality, and Interaction*.
- Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (pp. 1532–1543). ACL.
- Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2018). Automatic detection of fake news. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 3391–3401). ACL.

- Popat, K., Mukherjee, S., Strötgen, J., & Weikum, G. (2017). Where the truth lies: Explaining the credibility of emerging claims on the web and social media. In *WWW'17 Companion* (pp. 1003–1012). ACM.
- Popat, K., Mukherjee, S., Yates, A., & Weikum, G. (2018). 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* (pp. 22–32). ACL.
- Poria, S., Gelbukh, A., Hussain, A., Howard, N., Das, D., & Bandyopadhyay, S. (2013). Enhanced SenticNet with affective labels for concept-based opinion mining. *IEEE Intelligent Systems*, 28(2), 31–38.
- Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2018). A stylometric inquiry into hyperpartisan and fake news. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 231–240). ACL.
- Rangel, F., Giachanou, A., Ghanem, B., & Rosso, P. (2020). Overview of the 8th author profiling task at PAN 2020: Profiling fake news spreaders on Twitter. In L. Cappellato, C. Eickhoff, N. Ferro, & A. Névóel (Eds.), *CLEF 2020 Labs and Workshops, Notebook Papers*. CEUR Workshop Proceedings.
- Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 2931–2937). ACL.
- Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797–806). ACM.
- Severyn, A., & Moschitti, A. (2015). Twitter sentiment analysis with deep convolutional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 959–962). ACM.
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *SIGKDD Explorations Newsletter*, 19(1), 22–36.
- Shu, K., Wang, S., & Liu, H. (2017). Exploiting tri-relationship for fake news detection. *arXiv preprint arXiv:1712.07709*.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2), 267–307.
- Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro, L. (2017). Some like it hoax: Automated fake news detection in social networks. In *Proceedings of the 2nd Workshop on Data Science for Social Good* (pp. 1–15).
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Thorne, J., Vlachos, A., Cocarascu, O., Christodoulopoulos, C., & Mittal, A. (2018). The fact extraction and verification (FEVER) shared task. In *Proceedings of the 1st Workshop on Fact Extraction and Verification* (pp. 1–9).
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.
- Wang, W. Y. (2017). “liar, liar pants on fire”: A new benchmark dataset for fake news detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 422–426). ACL.
- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253.
- Zheng, Y., Mobasher, B., & Burke, R. (2016). Emotions in context-aware recommender systems. In *Emotions and Personality in Personalized Services: Models, Evaluation and Applications* (pp. 311–326).
- Zhou, X., Jain, A., Phoha, V. V., & Zafarani, R. (2020). Fake news early detection: A theory-driven model. *Digital Threats: Research and Practice*, 1(2), 1–25.
- Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. (2018). Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys*, 51(2), 32:1–32:36.

How to cite this article: Giachanou A, Rosso P, Crestani F. The impact of emotional signals on credibility assessment. *J Assoc Inf Sci Technol*. 2021;72:1117–1132. <https://doi.org/10.1002/asi.24480>