An Emotional Analysis of False Information in Social Media and News Articles

BILAL GHANEM, Universitat Politècnica de València, Spain
PAOLO ROSSO, Universitat Politècnica de València, Spain
FRANCISCO RANGEL, Symanto Research, Germany

Fake news is risky since it has been created to manipulate the readers' opinions and beliefs. In this work, we compared the language of false news to the real one of real news from an emotional perspective, considering a set of false information types (propaganda, hoax, clickbait, and satire) from social media and online news articles sources. Our experiments showed that false information has different emotional patterns in each of its types, and emotions play a key role in deceiving the reader. Based on that, we proposed a LSTM neural network model that is emotionally-infused to detect false news.

CCS Concepts: • **Computing methodologies** → *Neural networks*; **Natural language processing**.

Additional Key Words and Phrases: Fake News, Suspicious News, False information, Emotional Analysis

1 INTRODUCTION

With the complicated political and economic situations in many countries, some agendas are publishing suspicious news to affect public opinions regarding specific issues [23]. The spreading of this phenomenon is increasing recently with the large usage of social media and online news sources. Many anonymous accounts in social media platforms start to appear, as well as new online news agencies without presenting a clear identity of the owner. Twitter has recently detected a campaign[1] organized by agencies from two different countries to affect the results of the last U.S. presidential elections of 2016. The initial disclosures by Twitter have included 3,841 accounts. A similar attempt was done by Facebook, as they detected coordinated efforts to influence U.S. politics ahead of the 2018 midterm elections[2].

False information is categorized into 8 types[3] according to [39]. Some of these types are intentional to deceive where others are not. In this work, we are interested in analyzing 4 main types, i.e. **hoaxes**, **propagandas**, **clickbaits**, and **satires**. These types can be classified into two main categories -

[1]https://blog.twitter.com/official/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html

[2]https://www.businessinsider.es/facebook-coordinated-effort-influence-2018-us-midterm-elections-2018-7

[3]Types of False information: Fabricated, Propaganda, Conspiracy Theories, Hoaxes, Biased or one-sided, Rumors, Clickbait, and Satire News.

Authors' addresses: Bilal Ghanem, Universitat Politècnica de València, Valencia, Spain, bigha@doctor.upv.es; Paolo Rosso, Universitat Politècnica de València, Valencia, Spain, prosso@dsic.upv.es; Francisco Rangel, Symanto Research, Nürnberg, Germany, kico.rangel@gmail.com.

misinformation and disinformation - where misinformation considers false information that is published without the intent to deceive (e.g. satire). Disinformation can be seen as a specific kind of false information with the aim to mislead the reader (e.g. hoax, propaganda, and clickbait). **Propagandas** are fabricated stories spread to harm the interest of a particular party. **Hoaxes** are similar to propagandas but the main aim of the writer is not to manipulate the readers' opinions but to convince them of the validity of a paranoia-fueled story [32]. **Clickbait** is another type of disinformation that refers to the deliberate use of misleading headlines, thumbnails, or stories' snippets to redirect attention (for traffic attention). **Satire** is the only type of misinformation, where the writer's main purpose is not to mislead the reader, but rather to deliver the story in an ironic way (to entertain or to be sarcastic).

The topic of fake news is gaining attention due to its risky consequences. A vast set of campaigns has been organized to tackle fake news. The owner of Wikipedia encyclopedia created the news site WikiTribune[4] to encourage the evidence-based journalism.

Another way of addressing this issue is by fact-checking websites. These websites like *politifact.com*, *snopes.com* and *factchecking.org* aim to debunk false news by manually assess the credibility of claims that have been circulated massively in online platforms. These campaigns were not limited to the English language where other languages such as Arabic have been targeted by some sites like *fatabyyano.net*[5].

Hypothesis. Trusted news is recounting its content in a naturalistic way without attempting to affect the opinion of the reader. On the other hand, false news is taking advantage of the presented issue sensitivity to affect the readers' emotions which sequentially may affect their opinions as well. A set of works has been done previously to investigate the language of false information. The authors in [37] have studied rumours in Twitter. They have investigated a corpus of true and false tweets rumours from different aspects. From an emotional point of view, they found that false rumours inspired fear, disgust, and surprise in their replies while the true ones inspired joy and anticipation. Some kinds of false information are similar to other language phenomena. For example, satire by its definition showed similarity with irony language. The work in [12] showed that affective features work well in the detection of irony. In addition, they confirmed that positive words are more relevant for identifying sarcasm and negative words for irony [38]. The results of these works motivate us to investigate the impact of emotions on false news types. These are the research questions we aim to answer:

RQ1 *Can emotional features help detecting false information?*
RQ2 *Do the emotions have similar importance distributions in both Twitter and news articles sources?*
RQ3 *Which of the emotions have a statistically significant difference between false information and truthful ones?*
RQ4 *What are the top-N emotions that discriminate false information types in both textual sources?*

In this work, we investigate suspicious news in two different sources: Twitter and online news articles. Concerning the news articles source, we focus on the beginning part of them, since they are fairly long, and the emotional analysis could be biased by their length. We believe that the beginning part of false news articles can present a unique emotional pattern for each false information type since the writer in this part is normally trying to trigger some emotions in the reader.

Throughout the emotional analysis, we go beyond the superficial analysis of words. We hope that our findings in this work will contribute to fake news detection.

The key contributions of this article are:

[4]https://www.wikitribune.com
[5]fatabyyano is an Arabic term which means "to make sure".

- **Model**: We propose an approach that combines emotional information from documents in a deep neural network. We compare the obtained results with a set of baselines. The results show that our approach is promising.
- **Analysis**: We show a comprehensive analysis on two false information datasets collected from social media and online news articles, based on a large set of emotions. We compare the differences from an affective perspective in both sources, and obtain valuable insights on how emotions can contribute to detect false news.

The rest of the paper is structured as follows; After a brief review of related work in Section 2, Section 3 introduces our emotionally-infused model. Then, we present the evaluation framework in Section 4. Section 5 reports the experiments and the results, followed by an analysis on the false information types from emotional perspective in Section 6. Finally, the conclusions of this work are summarized in Section 7.

2 RELATED WORK

The work that has been done previously on the analysis of false information is rather small regarding the approaches that were proposed. In this section, we present some recent works on the language analysis and detection of false information.

Recent attempts tried to analyze the language of false news to give a better understanding. A work done in [36] has studied the false information in Twitter from a linguistic perspective. The authors found that real tweets contain significantly fewer bias markers, hedges, subjective terms, and less harmful words. They also found that propaganda news targets morals more than satires and hoaxes but less than clickbaits. Furthermore, satirical news contains more loyalty and fewer betrayal morals compared to propaganda. In addition, they built a model that combined a set of features (graph-based, cues words, and syntax) and achieved a good performance comparing to other baselines (71% vs. 59% macro-F1). Another similar work [32] has been done to characterize the language of false information (propaganda, hoax, and satire) in online news articles. The authors have studied the language from different perspectives: the existence of weak and strong subjectivity, hedges, and the degree of dramatization using a lexicon from Wiktionary. As well, they employed in their study the LIWC dictionary to exploit the existence of personal pronouns, swear, sexual, etc. words. The results showed that false news types tend to use first and second personal pronouns more than truthful news. Moreover, the results showed that false news generally uses words to exaggerate (subjectives, superlatives, and modal adverbs), and specifically, the satire type uses more adverbs. Hoax stories tend to use fewer superlatives and comparatives, and propagandas use relatively more assertive verbs. Moving away from these previous false information types, the work in [37] has focused on analyzing rumours in Twitter (from factuality perspective: True or False). They analyzed about 126,000 rumours and found that falsehood widespread significantly further, faster, deeper, and more broadly than truth in many domains. In addition, they found that false rumours are more novel than truthful ones, which made people more likely to share them. From an emotional perspective, they found that false rumours triggered "fear", "disgust", and "surprise" in replies while truthful ones triggered "anticipation", "sadness", "joy", and "trust". Another work [17] has studied the problem of detecting hoaxes by analyzing features related to the content in Wikipedia. The work showed that some features like hoaxes articles' length as well as the ratio of wiki markups (images, references, links to other articles and to external URLs, etc.) are important to discriminate hoaxes from legitimate articles. Many approaches have been proposed on fake news detection. In general, they are divided into social media and news claims-based approaches. The authors in [16, 20, 29, 34, 40] have proposed supervised methods using recurrent neural networks or by extracting manual features like a set of regular expressions, content-based, network-based

etc. As an example, the work by [3] assessed the credibility of tweets by analyzing trending topics. They used message-based, user-based, and propagation-based features, and they found that some features related to the user information like user's age, number of followers, statuse counts etc. have helped the most to discriminate truthful from deceitful tweets. Other news claims-based approaches [9, 14, 18, 26, 27] have been mainly focusing on inferring the credibility of the claims by retrieving evidences from Google or Bing search engines. These approaches have employed a different set of features starting from manual features (e.g. cosine similarity between the claims and the results, Alexa Rank of the evidence source, etc.) to a fully automatic approach using deep learning networks. A recent trend started to appear and is trying to approach the detection of fake news from a stance perspective. The aim is to predict how other articles orient to a specific fact [2, 10, 11].

3 EMOTIONALLY-INFUSED MODEL

In this section we describe the Emotionally-Infused Network we propose (EIN).

3.1 Emotional Lexicons

Several emotional models well-grounded in psychology science have been proposed, such as the ones by Magda Arnold [1], Paul Ekman [6], Robert Plutchik [25], and Gerrod Parrot [24]. On the basis of each of them, many emotional resources (lexicons) were built in the literature. In this work, we consider several emotional resources to increase the coverage of the emotional words in texts as well to have a wider range of emotions in the analysis. Concretely, we use EmoSenticNet, EmoLex, SentiSense, LIWC and Empath:

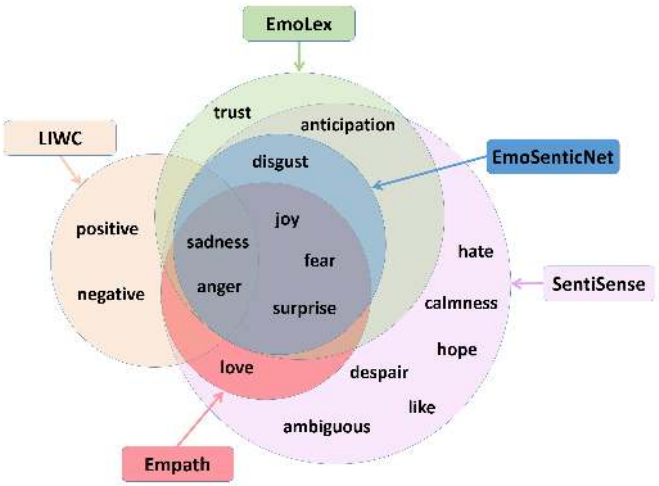

Fig. 1. The emotional lexicons with their own emotions.

- EmoSenticNet [28] is a lexical resource that assigns WordNet-Affect[6] emotion labels to SenticNet[7] concepts. It has a total of 13,189 entries annotated using the six Ekman's basic emotions.

[6]http://wndomains.fbk.eu/wnaffect.html
[7]https://sentic.net/

- EmoLex [22] is a word-emotion association lexicon that is labeled using the eight Plutchik's emotions. This lexicon contains 14,181 words.
- SentiSense [5] is a concept-based affective lexicon that attaches emotional meanings to concepts from the WordNet[8] lexical database. SentiSense has 5,496 words labeled with emotions from a set of 14 emotional categories, which is an edited version of the merge between Arnold, Plutchik, and Parrott models.
- LIWC [35] is a linguistic dictionary that contains 4,500 words categorized to analyze psycholinguistic patterns in text. Linguistic Inquiry and Word Count (LIWC) has 4 emotional categories: "sadness", "anger", "positive emotion", and "negative emotion".
- Empath [7] is a tool that uses deep learning and word embeddings to build a semantically meaningful lexicon for concepts. Empath uses Parrott's model for the emotional representation, but we use only the primary emotions (6 emotions) in the Pattrott's hierarchy ("love", "joy", "surprise", "anger", "sadness", "fear").

In our study we consider the 17 emotions that we shown in Figure 1[9]

3.2 Model

We choose an Long short-term memory (LSTM) [13] that takes the sequence of words as input and predicts the false information type. The input of our network is based on word embedding (content-based) and emotional features (see Figure 2).

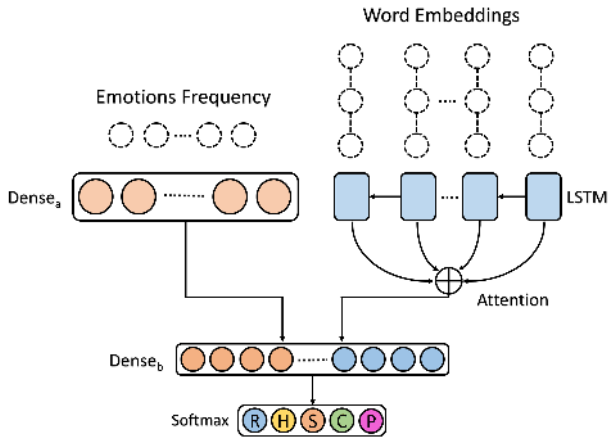

Fig. 2. Emotionally-infused neural network architecture for false information detection. RHSCP in the Softmax layer stands for Real, Hoax, Satire, Clickbait, and Propaganda respectively.

3.3 Input Representation

Our network consists of two branches. In the content-based one, we use an embedding layer followed by a LSTM layer. Then, we add an attention layer [30] to make this branch focus on (highlighting) particular words over others [10]. The attention mechanism assigns a weight to each word vector result from the LSTM layer with a focus on the classification class. The input representation for

[8]https://wordnet.princeton.edu

[9]We investigated the performance of different combinations of lexicons; in the article we show the results obtained with the best performing combination.

[10]We tested our model without the attention layer, but we got a lower result.

this branch is represented as follows: the input sentence S of length n is represented as $[S_1, S_2..S_n]$ where $S_n \in \mathbb{R}^d$; $\mathbb{R}^d$ is a d-dimensional word embedding vector of the i-th word in the input sentence. The output vectors of the words are passed to the LSTM layer, where the LSTM learns the hidden state h_t by capturing the previous timesteps (past features). The produced hidden state h_t at each time step is passed to the attention layer which computes a "context" vector c_t as the weighted mean of the state sequence h by:

$$c_t = \sum_{t=1}^{T} \alpha_t h_t, \tag{1}$$

Where T is the total number of timesteps in the input sequence and α_t is a weight computed at each time step t for each state h_t. This output vector is then concatenated with the output from the dense_a (see Figure 2) layer and passed to the dense_b layer, which precedes a final Softmax function to predict the output classes. Since the content-based branch is concatenated with the other emotional-based branch.

On the other hand, the input representation for the emotional-based branch is defined as follows: each lexicon is represented as L_{nm} where n is the number of emotional lexicons ($n \in [1, 5]$), and m is the number of emotions categories used depending on the emotion model (e.g. Plutchik, Arnold, etc.). In our implementation, the emotional vector of a Lexicon L_{nm} is built using word frequency and normalized by the input sentence's length. Each input sentence is represented using Equation 2.

$$v = L_{1m} \oplus L_{2m} \oplus L_{3m} \oplus L_{4m} \oplus L_{5m}, \tag{2}$$

Where $v \in \mathbb{R}^q$, $\oplus$ denotes the concatenate operation, and q is defined in Equation 3.

$$q = \sum_{i=1}^{n} ||L_i E_M||, \tag{3}$$

Next, the built emotion vector is fed to a dense layer to obtain emotion-specific representations of each input document (Equation 4).

$$a = f(W_a v + b_a), \tag{4}$$

where W_a and b_a are the corresponding weight matrix and bias terms, and f is an activation function, such as *ReLU*, *tanh*, etc.

4 EVALUATION FRAMEWORK

4.1 Datasets

Annotated data is a crucial source of information to analyze false information. Current status of previous works lacks available datasets of false information, where the majority of the works focus on annotating datasets from a factuality perspective. However, to analyze the existence of emotions across different sources of news, we rely on two publicly available datasets and a list contains suspicious Twitter accounts.

News Articles. Our dataset source of news articles is described in [32]. This dataset was built from two different sources, for the trusted news (real news) they sampled news articles from the English Gigaword corpus. For the false news, they collected articles from seven different unreliable news sites. These news articles include satires, hoaxes, and propagandas but not clickbaits. Since we are interested also in analyzing clickbaits, we slice a sample from an available clickbait dataset called Stop_Clickbait [4] that was originally collected from two sources: Wikinews articles' headlines and other online sites that are known to publish clickbaits. The satire, hoax, and propaganda news

Table 1. News articles and Twitter datasets' statistics.

Category	News Articles	Twitter
Satire	5,750 (18%)	12,502 (8%)
Hoax	5,750 (18%)	6,247 (4%)
Propaganda	5,750 (18%)	66,225 (43.5%)
Clickbait	5,750 (18%)	36,103 (23.5%)
Real News	8,550 (28%)	30,949 (21%)
Total	**31,550**	**152,026**

articles are considerably long (some of them reach the length of 5,000 words). This length could affect the quality of the analysis as we mentioned before. We focus on analyzing the initial part of the article. Our intuition is that it is where emotion-bearing words will be more frequent. Therefore, we shorten long news articles into a maximum length of N words (N=300). We choose the value of N based on the length of the shortest articles. Moreover, we process the dataset by removing very short articles, redundant articles or articles that do not have a textual content[11].

Twitter. For this dataset, we rely on a list of several Twitter accounts for each type of false information from [36]. This list was created based on public resources that annotated suspicious Twitter accounts. The authors in [36] have built a dataset by collecting tweets from these accounts and they made it available. For the real news, we merge this list with another 32 Twitter accounts from [15]. In this work we could not use the previous dataset[12] and we decide to collect tweets again. For each of these accounts, we collected the last M tweets posted (M=1000). By investigating these accounts manually, we found that many tweets just contain links without textual news. Therefore, to ensure of the quality of the crawled data, we chose a high value for M (also to have enough data). After the collecting process, we processed these tweets by removing duplicated, very short tweets, and tweets without textual content. Table 1 shows a summary for both datasets.

4.2 Baselines

Emotions have been used in many natural language processing tasks and they showed their efficiency [31]. We aim at investigating their efficiency to detect false information. In addition to EIN, we created a model (Emotion-based Model) that uses emotional features only by converting the input documents into vectors of emotions frequency (see Equation 2) and compare it to two baselines. Our aim is to investigate if the emotional features independently can detect false news. The two baselines of this model are Majority Class baseline (MC) and the Random selection baseline (RAN).

For the EIN model, we compare it to different baselines: **a**) The first one is bag-of-words with a support vector machine classifier (BOW-SVM). We test different classifiers, and we choose SVM since it gives the highest result in the 10-fold Cross Validation (CV); **b**) We use another baseline that is based on word embeddings where for each input document we extract an average word embedding vector by taking the mean of the embeddings for the document's words. Similarly, we test different classifiers and the Logistic Regression classifier shows the best performance (WE-LR); **c**) The last baseline is the same as our neural architecture but without the emotional features branch: an LSTM layer followed by attention and dense layers.

[11]e.g. "BUYING RATES US 31.170 .."

[12]Due to Twitter terms of usage, the authors provided in their dataset the ids of the tweets and when we tried to collect these tweets many of them were deleted.

Table 2. The results of the Emotion-based Model with the emotional features comparing to the baselines.

	Accuracy	Macro-Precision	Macro-Recall	Macro-F1
News Articles				
Majority Class	0.34	0.07	0.20	0.10
Random Selection	0.21	0.21	0.21	0.20
Emotion-based Model	0.50	0.48	0.51	0.48
Twitter				
Majority Class	0.44	0.09	0.20	0.12
Random Selection	0.20	0.20	0.20	0.18
Emotion-based Model	0.52	0.55	0.38	0.41

5 EXPERIMENTS AND RESULTS

5.1 Emotion-based Model

In our experiments, we use 20% of each of the datasets for testing and we apply 10-fold cross-validation on the remain part for selecting the best classifier as well for tuning it. We tested many classifiers and we finally choose Random Forest for both datasets since it obtained the best results[13]. Table 2 presents the classification results on both datasets.

The results in both datasets show that emotional features clearly detect false news, compared to the baselines (**RQ1**). The emotional features perform better in the news articles dataset compared with these of tweets. We are interested in investigating also how good are the emotional features in detecting each class comparing to the RAN baseline. We choose the RAN baseline since it shows better results with regard to macro-F1 score. For doing so, we investigated the True Positive (TP) classification ratio for each class in each dataset.

The clickbait class shows the highest TPs comparing to the other classes. From this we can infer that clickbaits exploit emotions much more than the other classes to deceive the reader. It is worth to mention that for the hoax class the proposed approach is better than the random baselines with a small ratio (4% difference). This could be justified by the fact that hoaxes, by definition, try to convince the reader of the credibility of a false story. Hence, the writer tries to deliver the story in a normal way without allowing the reader to fall under suspicion. The number of instances related to the false information classes in the news articles dataset is the same. Therefore, there is not a majority class that the classifier can be biased to. This is not the case in the Twitter dataset. For the Twitter dataset, the dataset is not balanced. Therefore, where the results are biased by the majority class (propaganda). But in general, all the classes' TP ratios are larger than the corresponding ones obtained with RAN baseline. From these results, we can conclude that suspicious news exploits emotions with the aim to mislead the reader. Following, we present the results obtained by the proposed emotionally-infused model.

5.2 Emotionally-Infused Model

In the neural model, to reduce the computational costs, instead of the cross-validation process we take another 20% from the training part as a validation set[14] (other than the 20% that is prepared for

[13]The other classifiers that we tested are: Support vector machine (testing both kernels), naive bayes, logistic regression, k-nearest neighbor and multilayer perceptron.

[14]We are forced to use different validation scenarios because for selecting the best parameters in the classical machine learning Scikit-Learn library we used Grid Search technique where CV is the only option for tuning. On the other hand, it is too expensive computationally to use CV to tune a deep neural network using a large parameter space.

Table 3. Models' parameters used in the three datasets (News articles, Twitter, Stop_Clickbaits). LSTM: the 3rd baseline, EIN: Emotionally-Infused Network.

Parameter	News Articles		Twitter		Stop_Clickbait	
	LSTM	EIN	LSTM	EIN	LSTM	EIN
LSTM units	140	90	180	180	120	120
$Dense_a$ units	-	320	-	100	-	60
$Dense_b$ units	320	60	120	60	260	120
Batch	64	64	64	64	32	32
Activation	relu	relu	relu	relu	tanh	relu
Optimizer	adadelta	adam	adadelta	rmsprop	rmsprop	Adam
$Drop_c$	0.5	0.5	0.5	0.2	0.2	0.2
$Drop_d$	0.2	0.1	0.2	0.2	0.2	0.2

testing). For the pretrained word embeddings, we use Google News Word2Vec 300-Embeddings[15] in the neural network as well as in the W2V-LR baseline. For the classical machine learning classifiers for the baselines, we use the Scikit-Learn python library, and for the deep learning network, we use Keras library with Tensorflow as backend. To tune our deep learning network (hyper-parameters), we use the Hyperopt[16] library. And to reduce the effect of overfitting, we use early stopping technique.

In Table 3 we summarize the parameters with respect to each dataset. We have to mention that we use Dropout after the dense layer in the emotional features branch ($Drop_c$) as well as after the attention layer in the other one ($Drop_d$) before the concatenation process. Since it is a multiclass classification process, we use categorical cross-entropy loss function. A summary of the models' parameters is presented in Table 3.

Table 4 summarizes the performance of the proposed model in comparison to those obtained by the baselines. We report Macro- precision, recall, and F1, including also the metric of accuracy; for comparing the models' results we consider the macro of metrics since it shows an averaged result over all the classes. The baselines that we propose clearly show high results, where the LSTM baseline has the best performance in news articles dataset. In Twitter there is a different scenario, the BOW-SVM baseline shows a higher performance with respect to LSTM. We are interested in investigating the reason behind that. Therefore, we checked the coverage ratio of the used embeddings in the Twitter dataset. We have to mention that we excluded stop words during representing the input documents using the pre-trained Google News word embeddings[17]. In the news articles dataset, we found that the coverage ratio of the embeddings is around 94% while in Twitter it is around 70%. Therefore, we tuned the word embeddings during the training process to improve the document's representation since we have a larger dataset from Twitter. This process contributed with 1.9% on the final macro-F1 results in Twitter (the result without tuning is 0.54). Even though, the results obtained with the LSTM baseline is still lower than the one obtained with BOW-SVM. This experiment gives us some intuition that the weaker performance on Twitter may be due to the embeddings. Therefore, we tried different embeddings but none of them improved the result[18]. The second baseline (W2V-LR) proved the same issue regarding the embeddings. The

[15]https://code.google.com/archive/p/word2vec/

[16]https://github.com/hyperopt/hyperopt

[17]The existence of stop words is importance to conserve the context in the LSTM network, but we got better results without them.

[18]e.g. Glove (using multiple embedding dimensions) and FastText.

Table 4. Results of the proposed model (EIN) vs. the baselines.

	Accuracy	Macro-Precision	Macro-Recall	Macro-F1
News Articles				
BOW+SVM	0.74	0.72	0.71	0.71
W2V+LR	0.72	0.70	0.70	0.70
LSTM	0.75	0.77	0.74	0.74
EIN	0.80	0.79	0.80	0.79
Twitter				
BOW+SVM	0.63	0.60	0.56	0.57
W2V+LR	0.53	0.49	0.35	0.36
LSTM	0.64	0.65	0.54	0.56
EIN	0.65	0.61	0.59	0.60

Table 5. F1 score results of the proposed model (EIN) vs. the baselines with respect to each class.

	clickbait	hoax	propaganda	realnews	satire
News Articles					
BOW+SVM	0.90	0.54	0.59	0.84	0.66
W2V+LR	0.85	0.57	0.66	0.83	0.57
LSTM	0.88	0.67	0.65	0.80	0.68
EIN	0.91	0.69	0.75	0.85	0.76
Twitter					
BOW+SVM	0.49	0.40	0.70	0.67	0.62
W2V+LR	0.32	0.07	0.66	0.45	0.32
LSTM	0.48	0.35	0.72	0.65	0.63
EIN	0.53	0.37	0.74	0.67	0.67

W2V-LR macro-F1 result in the news articles dataset is competitive, where it is much lower in Twitter. The usage of LSTM is two folds: in addition to being a good baseline, it shows also how much the emotional features contribute in the emotionally-infused network.

EIN results outperform the baselines with a large margin (around 3% in Twitter and 6% in news articles), especially in the news articles dataset. The margin between EIN and the best baseline is lower in the Twitter dataset. The results also show that combining emotional features clearly boosts the performance. We can figure out the improvement by comparing the results of EIN to LSTM. EIN shows superior results in news articles dataset with regard to the LSTM (0.79). A similar case appears in the Twitter dataset but with a lower margin (0.60). The results of EIN in Twitter dataset show that emotional features help the weak coverage of word embeddings to improve the performance as well as to overcome the BOW-SVM baseline.

Furthermore, to investigate the improvement that the emotions produced in the detection of the classes, in Table 5 we present the F1 score results for each class. For the news articles dataset, the results show that employing emotions in the EIN model improves the detection in all the cases, especially in real news, propaganda, and satire classes. On the other hand, for the Twitter dataset, emotions contribute especially in the case of clickbait and satire.

We observed before that clickbait TP's ratio of the news articles dataset is the highest one, and this result points out that the clickbait class is less difficult to detect specifically from an

emotional perspective. Therefore, in order to assess how our model separates false information types, we employ dimensionality reduction using t-distributed Stochastic Neighbor Embedding (T-SNE) technique [21] to project the document's representation from a high dimensional space to a 2D plane. Thus, we project the embeddings in EIN by extracting them from the outputs of Dense_b layer (see Figure 3). We extract the embeddings twice, once from a random epoch (epoch 10) at the beginning of the training phase and the other at the last epoch.

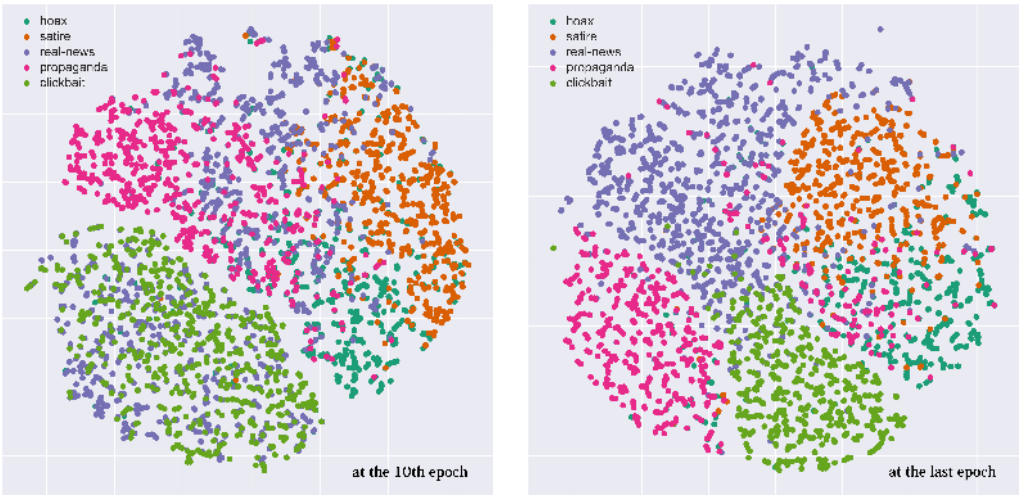

Fig. 3. Projection of documents representation from the news articles dataset.

Our aim from the early epoch projection is to validate what we have noticed: the clickbait class is less difficult to detect with regard to the other classes. As we can notice in the 10-epoch plot, the clickbait class needs few epochs to be separated from the other types, and this supports what we found previously in the manual investigation of the classes' TP ratios. Despite this clear separation, there is still an overlapping with some real-news records. This results points out that emotions in clickbaits play a key role in deceiving the reader. Also, the figure shows that the disinformation classes still need more training epochs for better separation. Real-news records are totally overlapped with the false information classes as well as the false information classes with each other. On the other hand, for the last epoch, clearly, the classes are separated from each other and the more important, from the real news. But generally, there still a small overlapping between satires and hoaxes as well few records from the propaganda class.

5.3 EIN as Clickbaits Detector

From the previous results in Section 5.1 as well as from what we notice in Figure 3, EIN obtains a clear separability of the clickbait class. These observations motivate us to investigate EIN as clickbait detector. Concretely, we test EIN on the source of our clickbait instances [4] in the news articles dataset. As we mentioned previously, this dataset originally was built using two different text sources. For clickbaits, the authors have manually identified a set of online sites that publish many clickbait articles. Whereas for the negative class, they collected headlines from a corpus of Wikinews articles collected in other research work. They took 7,500 samples from each class for the final version of the dataset. The authors also proposed a clickbaits detector model[19] that employed

[19]We use Stop_Clickbait to refer to this approach in the rest of the experiments.

Table 6. The performance of EIN on the clickbaits dataset using 10-fold CV.

	Accuracy	Precision	Recall	F1
Stop_Clickbait	0.93	0.95	0.90	0.93
LSTM	95	0.95	0.96	0.95
EIN	0.96	0.96	0.97	0.96

a combination of features: *sentence structure* (sentence length, average length of words, the ratio of the number of stop words to the number of thematic words and the longest separation between the syntactically dependent words), *word patterns* (presence of cardinal number at the beginning of the sentence, presence of unusual punctuation patterns), *clickbait language* (presence of hyperbolic words, common clickbait phrases, internet slangs and determiners), and *N-grams features* (word, Part-Of-Speech, and syntactic n-grams). Using this set of features group, the authors tested different classifiers where SVM showed the state-of-the-art results. They considered Accuracy, Precision, Recall and F1 to compare their approach to a baseline (an online web browser extension for clickbaits detection called Downworthy[20]).

In this experiment, we consider the third baseline (LSTM) to observe the improvement of the emotional features in the EIN model. Different from the previous experiments, this is a binary classification task. Therefore, we use binary cross-entropy as loss function and we change the Softmax layer to a Sigmoid function. The new parameters for both LSTM and EIN models are mentioned in Table 3.

In Table 6 we present the results of the Stop_Clickbait approach, LSTM baseline, and the EIN model. The results show that our baseline outperforms the proposed clickbait detector with a good margin. Furthermore, the results of the EIN are superior to the LSTM and the Stop_Clickbait detector. Considering emotions in the EIN deep learning approach improved the detection of false information. This is due to the fact that in clickbaits emotions are employed to deceive the reader.

6 DISCUSSION

The results show that the detection of suspicious news in Twitter is harder than detecting them in news articles. Overall, the results of EIN showed that emotional features improve the performance of our model, especially in the case of the news articles dataset. We manually inspected the Twitter dataset and observed that the language of the tweets has differences compared to the news articles one. We found that news in Twitter has many abbreviations (amp, wrt, JFK...etc.), bad words abbreviations (WTF, LMFO...etc.), informal language presentation, and typos. This reduces the coverage ratio of word embeddings. We also noticed that suspicious news in Twitter are more related to sexual issues. To validate our observations, we extracted the mean value of sexual words using a list of sexual terms [8]. The mean value is the average number of times a sexual/bad word appears in a tweet normalized by the length of the tweet. The mean value in Twitter is 0.003[21] while in news articles is 0.0024. Similarly, suspicious news in Twitter presented more insulting words[22] than in news articles where the mean value in Twitter is 0.0027 and 0.0017 in news articles.

Following, we focus on analyzing false information from an emotional perspective. We are aiming to answer the rest of the questions, **RQ2**, **RQ3**, and **RQ4**.

RQ2 *Do the emotions have similar importance distributions in both Twitter and news articles sources?*

[20]http://downworthy.snipe.net/

[21]The mean value is normalized by the sentence length since the news articles documents are longer than Tweets.

[22]Insult-wiki: http://www.insult.wiki/wiki/Insult_List

Intuitively, the emotions contribution in the classification process is not the same, where some words could manifest the existence of specific kind of emotions rather than others. To investigate this point, we use Information Gain (IG) in order to identify the importance of emotions in discriminating between real and all the other types of false news (multiclass task) in both Twitter and news articles datasets (see Figure 4). Before going through the ranking of features importance, we notice that the emotions ranking shapes are very similar in both Twitter and news articles. This states that despite the fact that the language is different, both sources have similar overall emotions distribution. In other words, false news employs a similar emotional pattern in both text sources. Since the news language in Twitter is not presented clearly as in news articles, this observation can help to build a cross-source system that is trained on suspicious news from news articles to detect the corresponding ones in Twitter. Figure 4 shows also that the emotion "joy" is the most important emotion in both datasets. It also mentions that "despair" and "hate" are almost not used in the classification process. The ranking of the features in both sources is different, where in the news articles dataset the top important emotions are "joy", "anticipation", "fear", and "disgust" respectively. On the other hand, the top ones in Twitter are "joy", "sadness", "fear", and "disgust".

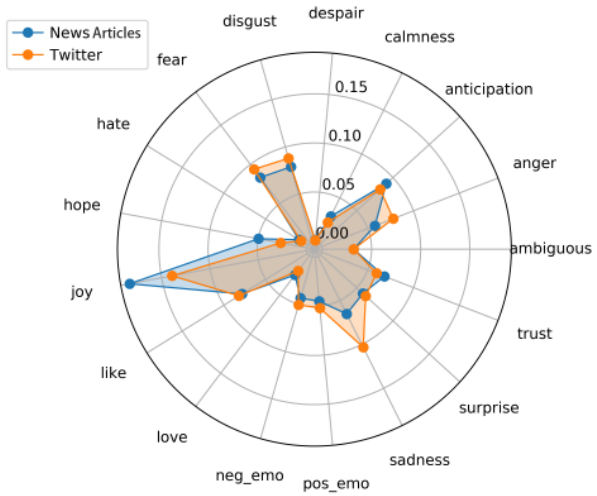

Fig. 4. Best ranked features according to Information Gain.

RQ3 *Which of the emotions have a statistically significant difference between false information and truthful ones?*

We measure statically significant differences using the t-test on emotions across real news and false news (binary task) in the both datasets in Figure 5. These findings provide a deeper understanding of the EIN performance. The results show that "joy", "neg_emo", "ambiguous", "anticipation", "calmness", "disgust", "trust" and "surprise" have significant statistical differences between real and suspicious news in both datasets. Some other emotions such as "despair" and "anger" have no statistical difference in both datasets. It turns out that the results we obtain are generally consistent with the IG results in research question **RQ2**. We notice in the IG analysis that some emotions have a higher importance in one of the news sources: "sadness", "anger", and "fear" have a higher importance in Twitter than in news articles, and the opposite for "hope". We observe the same findings using the t-test.

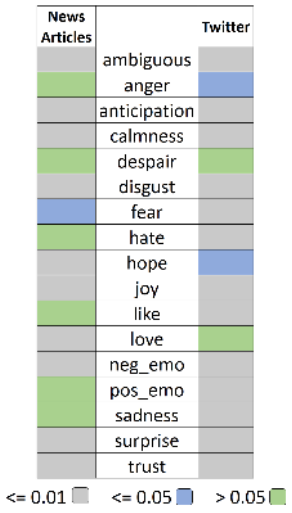

Fig. 5. Statistical significant differences between false and real news on Twitter and news articles datasets using t-test.

Table 7. The top 3 most important emotions in each false information type.

Rank	clickbait	hoax	propaganda	satire
News Articles				
1	surprise	hope	joy	disgust
2	neg_emo	anger	fear	neg_emo
3	like	like	calmness	pos_emo
Twitter				
1	surprise	like	fear	pos_emo
2	neg_emo	disgust	hope	disgust
3	fear	anticipation	calmness	sadness

RQ4 *What are the top-N emotions that discriminate false information types in both textual sources?*

False information types are different in the way they present the news to the reader. This raises a question: what are the top employed emotions in each type of false information? In Table 7, we present the first three[23] emotions that contribute mostly to the classification process to each type. This can indicate to us what are the emotion types that are used mostly in each type of false information.

Table 7 shows that clickbaits express "surprise" and "negative emotion" at the most. This validates the definition of clickbaits as "attention redirection" by exploiting the reader and convincing him/her that there is an unexpected thing with negative emotion. The result of seeing "fear" in the top features in Twitter is interesting; one of the recent studies is presenting the hypothesis that says: *curiosity is the best remedy for fear* [19] based on psychological interpretations. Taking into account the definition of clickbaits as "attention redirection", looking at our results, we can

[23]We used SVM classifier coefficients (linear kernel) to extract the most important emotions to each classification class.

proof this hypothesis. Furthermore, despite the language differences in both datasets, we obtain almost the same results, which emphasize our results. For hoaxes, it is not simple to interpret a specific pattern of emotions in the results. We might justify it by the fact that hoaxes are written to convince the reader of the validity of a story. Therefore, the writer is trying to present the story in a normal way (truthful) similar to a real story. Therefore, the top emotions are not unique to the hoax type. But what we find from the top hoaxes emotions in both datasets is that they are generally different except the emotion "like". Despite the natural narrative way of presenting the story, the analysis shows that the writer still uses "like" to grab reader's attention smoothly. Propaganda type has clearer emotional interpretation considering its definition. We find that propaganda expresses "joy", "fear" and at the same time "calmness" in the news articles. Both "joy" and "fear" are contrary from an emotional polar perspective, where "joy" shows the extreme of the positive emotions and "fear" the extreme negative, and at the same time, "calmness" is present. The emotional shifting between the two extremes is a clear attempt of opinion manipulation from an emotional perspective. We obtain a similar emotion set from Twitter, but instead of "joy" we get "hope". Lastly, satire is defined as a type of parody presented in a typical format of mainstream journalism, but in a similar way to irony and sarcasm phenomena [33]. The results of the analysis show that "disgust" and "positive emotion" are present in both datasets, but we get "negative emotion" in the news articles and "sadness" in Twitter (both are placed in the negative side of emotions). We are interested in investigating the cause of the emotion "disgust" which appeared in the results from both datasets. We conduct a manual analysis on the text of the satire type in both datasets in order to shed some light on the possible causes. We notice that the satire language in the news often employs the emotion "disgust" to give a sense of humor. Figure 6 shows some examples from the news articles dataset highlighting the words that triggered the emotion "disgust".

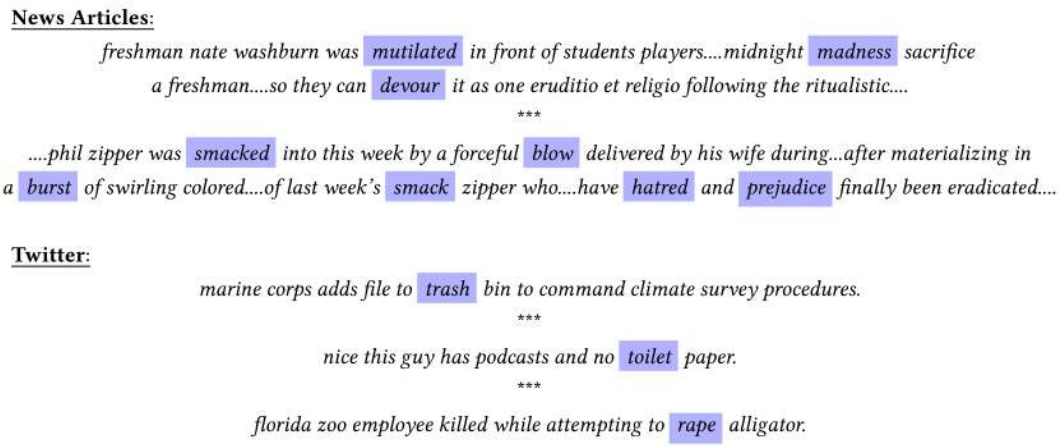

Fig. 6. Examples from news articles and Twitter datasets trigger the emotion "disgust".

7 CONCLUSIONS AND FUTURE WORK

In this article we have presented an emotionally-infused deep learning network that uses emotional features to identify false information in Twitter and news articles sources. We performed several experiments to investigate the effectiveness of the emotional features in identifying false information. We validated the performance of the model by comparing it to a LSTM network and other baselines. The results on the two datasets showed that clickbaits have a simpler manipulation

language where emotions help detecting them. This demonstrates that emotions play a key role in deceiving the reader. Based on this result, we investigated our model performance on a clickbaits dataset and we compared it to the state-of-the-art performance. Our model showed superior results near to 96% F1 value.

Overall results confirmed that emotional features have boosted EIN model performance achieving better results on 3 different datasets (**RQ1**). These results emphasized the importance of emotional features in the detection of false information. In Twitter, false news content is deliberately sexual oriented and it uses many insulting words. Our analysis showed that emotions can help detecting false information also in Twitter.

In the analysis section, we answered a set of questions regarding the emotions distribution in false news. We found that emotions have similar importance distribution in Twitter and news articles regardless of the differences in the used languages (**RQ2**). The analysis showed that most of the used emotions have statistical significant difference between real and false news (**RQ3**). Emotions plays a different role in each type of false information in line with its definition (**RQ4**). We found that clickbaits try to attract the attention of the reader by mainly employing the "surprise" emotion. Propagandas are manipulating the feelings of the readers by using extreme positive and negative emotions, with triggering a sense of "calmness" to confuse the readers and enforcing a feeling of confidence. Satire news instead use the "disgust" emotion to give a sense of humor. To sum up, we can say that the initial part of false news contains more emotions than the rest of document. Our approach exploit this fact for their detection.

To the best of our knowledge, this is the first work that analyzes the impact of emotions in the detection of false information considering both social media and news articles. As a future work, the results of our approach as a clickbaits detector motivate us to develop for a clickbaits detector as a web browser extension. Also, we will study how the emotions flow inside the articles of each kind of false information, which is worthy to be investigated as the results of this work confirmed.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Magda B Arnold. 1960. *Emotion and Personality.* Columbia University Press.

[2] Gaurav Bhatt, Aman Sharma, Shivam Sharma, Ankush Nagpal, Balasubramanian Raman, and Ankush Mittal. 2018. Combining Neural, Statistical and External Features for Fake News Stance Identification. In *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 1353–1357.

[3] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information Credibility on Twitter. In *Proceedings of the 20th international conference on World wide web*. ACM, 675–684.

[4] Abhijnan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. 2016. Stop Clickbait: Detecting and Preventing Clickbaits in Online News Media. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE Press, 9–16.

[5] Jorge Carrillo de Albornoz, Laura Plaza, and Pablo Gervás. 2012. SentiSense: An easily scalable concept-based affective lexicon for sentiment analysis.. In *LREC*. 3562–3567.

[6] Paul Ekman. 1992. An Argument for Basic Emotions. *Cognition & emotion* 6, 3-4 (1992), 169–200.

[7] Ethan Fast, Binbin Chen, and Michael S Bernstein. 2016. Empath: Understanding Topic Signals in Large-Scale Text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4647–4657.

[8] Simona Frenda, Bilal Ghanem, and Manuel Montes-y Gòmez. 2018. Exploration of Misogyny in Spanish and English Tweets. In *Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018)*, Vol. 2150. Ceur Workshop Proceedings, 260–267.

[9] Bilal Ghanem, Manuel Montes-y Gòmez, Francisco Rangel, and Paolo Rosso. 2018. UPV-INAOE-Autoritas - Check That: An Approach based on External Sources to Detect Claims Credibility. In *In Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, CLEF '18*.

[10] Bilal Ghanem, Paolo Rosso, and Francisco Rangel. 2018. Stance Detection in Fake News A Combined Feature Representation, In Proceedings of the First Workshop on Fact Extraction and VERification (FEVER). *EMNLP 2018*, 66–71.

[11] Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M Meyer, and Iryna Gurevych. 2018. A Retrospective Analysis of the Fake News Challenge Stance Detection Task. *arXiv preprint arXiv:1806.05180* (2018).

[12] Delia Irazú Hernańdez Farías, Viviana Patti, and Paolo Rosso. 2016. Irony Detection in Twitter: The Role of Affective Content. *ACM Transactions on Internet Technology (TOIT)* 16, 3 (2016), 19.

[13] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural computation* 9, 8 (1997), 1735–1780.

[14] Georgi Karadzhov, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, and Ivan Koychev. 2017. Fully Automated Fact Checking Using External Sources. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*. 344–353.

[15] Alireza Karduni, Ryan Wesslen, Sashank Santhanam, Isaac Cho, Svitlana Volkova, Dustin Arendt, Samira Shaikh, and Wenwen Dou. 2018. Can You Verifi This? Studying Uncertainty and Decision-Making About Misinformation Using Visual Analytics. (2018).

[16] Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task Learning for Rumour Verification. *arXiv preprint arXiv:1806.03713* (2018).

[17] Srijan Kumar, Robert West, and Jure Leskovec. 2016. Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes. In *Proceedings of the 25th international conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 591–602.

[18] Xian Li, Weiyi Meng, and Clement Yu. 2011. T-verifier: Verifying Truthfulness of Fact Statements. In *Data Engineering (ICDE), 2011 IEEE 27th International Conference on*. IEEE, 63–74.

[19] Mario Livio. 2017. *Why?: What Makes Us Curious*. Simon and Schuster Publishing.

[20] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting Rumors from Microblogs with Recurrent Neural Networks. In *IJCAI*. 3818–3824.

[21] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing Data Using t-SNE. *Journal of machine learning research* 9, Nov (2008), 2579–2605.

[22] Saif M Mohammad and Peter D Turney. 2010. 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*. Association for Computational Linguistics, 26–34.

[23] Brendan Nyhan and Jason Reifler. 2010. When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior* 32, 2 (2010), 303–330.

[24] W Gerrod Parrott. 2001. *Emotions in Social Psychology: Essential Readings*. Psychology Press.

[25] Robert Plutchik. 2001. The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist* 89, 4 (2001), 344–350.

[26] Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2016. Credibility Assessment of Textual Claims on the Web. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2173–2178.

[27] Kashyap Popat, Subhabrata Mukherjee, Andrew Yates, and Gerhard Weikum. 2018. DeClarE: Debunking Fake News and False Claims using Evidence-Aware Deep Learning. *arXiv preprint arXiv:1809.06416* (2018).

[28] Soujanya Poria, Alexander Gelbukh, Amir Hussain, Newton Howard, Dipankar Das, and Sivaji Bandyopadhyay. 2013. Enhanced SenticNet with Affective Labels for Concept-Based Opinion Mining. *IEEE Intelligent Systems* 28, 2 (2013), 31–38.

[29] Vahed Qazvinian, Emily Rosengren, Dragomir R Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying Misinformation in Microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 1589–1599.

[30] Colin Raffel and Daniel PW Ellis. 2015. Feed-Forward Networks with Attention Can Solve Some Long-Term Memory Problems. *arXiv preprint arXiv:1512.08756* (2015).

[31] Francisco Rangel and Paolo Rosso. 2016. On the Impact of Emotions on Author Profiling. *Information processing & management* 52, 1 (2016), 73–92.

[32] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 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*. 2931–2937.

[33] Victoria L Rubin, Yimin Chen, and Niall J Conroy. 2015. Deception Detection for News: Three Types of Fakes. In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*.

American Society for Information Science, 83.

[34] Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. Csi: A Hybrid Deep Model for Fake News Detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 797–806.

[35] Yla R Tausczik and James W Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.

[36] Svitlana Volkova, Kyle Shaffer, Jin Yea Jang, and Nathan Hodas. 2017. 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 (Volume 2: Short Papers)*, Vol. 2. 647–653.

[37] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The Spread of True and False News Online. *Science* 359, 6380 (2018), 1146–1151.

[38] Po-Ya Angela Wang. 2013. # Irony or# Sarcasm ĂŤ A Quantitative and Qualitative Study Based on Twitter. In *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27)*. 349–356.

[39] Savvas Zannettou, Michael Sirivianos, Jeremy Blackburn, and Nicolas Kourtellis. 2018. The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans. *arXiv preprint arXiv:1804.03461* (2018).

[40] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. 2015. Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1395–1405.