Leveraging Emotional Signals for Credibility Detection

Anastasia Giachanou
Universitat Politècnica de València
València, Spain
angia9@upv.es

Paolo Rosso
Universitat Politècnica de València
València, Spain
prosso@dsic.upv.es

Fabio Crestani
Università della Svizzera italiana
Lugano, Switzerland
fabio.crestani@usi.ch

ABSTRACT

The spread of false information on the Web is one of the main problems of our society. Automatic detection of fake news posts is a hard task since they are intentionally written to mislead the readers and to trigger intense emotions to them in an attempt to be disseminated in the social networks. Even though recent studies have explored different linguistic patterns of false claims, the role of emotional signals has not yet been explored. In this paper, we study the role of emotional signals in fake news detection. In particular, we propose an LSTM model that incorporates emotional signals extracted from the text of the claims to differentiate between credible and non-credible ones. Experiments on real world datasets show the importance of emotional signals for credibility assessment.

CCS CONCEPTS

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

KEYWORDS

fake news detection, credibility detection, emotional signals, LSTM

ACM Reference Format:
Anastasia Giachanou, Paolo Rosso, and Fabio Crestani. 2019. Leveraging Emotional Signals for Credibility Detection. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3331184.3331285

1 INTRODUCTION

Recent years have seen a large increase in the amount of fake news posted online. The spread of fake news has vast negative effects on society. For example, posts with false information, mainly propagated in social media, led to a decrease of *Measles, Mumps, & Rubella* (MMR) vaccination rates[1] causing in 2017 one of the worst measles outbreak in decades, a disease that was almost eradicated. Automatic fake news detection is very challenging since fake news is a mixture of false and true information, and intentionally written to confuse readers and to trigger intense emotions to them. To help humans detect fake news, several fact-checking websites (e.g., snopes.com, politifact.com) were developed. However, these websites require expert analysis that takes time.

Recent studies have tried to address the problem of fake news detection. The majority of the studies rely on textual information. Rashkin et al. [14] used linguistic information from the claims to address the credibility detection problem. On the other hand, Popat et al. [12] considered external evidence from articles that were retrieved from the Web and which were relevant to the claims.

One characteristic of fake news is that they are intentionally written to evoke emotions to the readers in an attempt to be believed and be disseminated in the social networks. One study that has explored emotions in rumours is presented by Vosoughi et al. [16]. Vosoughi et al. investigated true and false rumours on Twitter and found that false rumours triggered fear, disgust and surprise in their replies, whereas the true rumours triggered joy, sadness, trust and anticipation. However, they did not explore the effectiveness of emotions in automatic false information detection.

Although emotions seem to play an important role in false information detection, none of the proposed systems have incorporated emotions to investigate their effectiveness. In this paper, we propose the EmoCred system which incorporates emotional signals into a Long Short Term Memory (LSTM) neural network to differentiate between credible and non-credible claims. We explore three different approaches for generating the emotional signals from the claims: (i) a lexicon-based approach that is based on the number of emotional words that appear in the claim, (ii) an approach that calculates the emotional intensity expressed in the claim using an emotional intensity lexicon and, (iii) a neural network that predicts the level of emotional intensity that can be triggered to the users. Experiments on real world datasets demonstrate the effectiveness of incorporating emotional signals to the systems for credibility assessment.

2 RELATED WORK

Deep learning approaches have recently gained tremendous popularity for different Natural Language Processing tasks. This can be attributed to their ability to capture information directly from the data. Recurrent Neural Networks (RNN) are designed to recognize patterns in sequences of data such as text, speech, or numerical times series data. LSTM networks [6] is a type of RNN network that has recently been applied for various problems such as fake news detection and sentiment analysis.

Fake news detection and credibility assessment have recently attracted a lot of research attention. Rashkin et al. [14] trained an LSTM model using linguistic features from the claims' text to distinguish between credible and non-credible claims. They incorporated various linguistic features extracted using the Linguistic Inquiry and Word Count (LIWC) dictionary [15] such as personal pronouns

[1] https://www.weforum.org/agenda/2017/08/scientists-can-vaccinate-against-the-post-truth-era

SIGIR '19, July 21–25, 2019, Paris, France

ACM ISBN 978-1-4503-6172-9/19/07…$15.00
https://doi.org/10.1145/3331184.3331285

and swear words. Wang [17] combined a neural network with different meta-data of the claims such as the author and the subject. Ma et al. [7] focused on detecting rumours in social media and proposed an RNN model for learning the hidden representations that can capture the variation of contextual information of relevant posts over time.

Other researchers proposed to use external evidence from the Web for credibility detection. Popat [11] combined different information extracted from external resources such as stance detection and trustworthiness of the sources. The different sources of information are combined via a pipeline of supervised classifiers. Popat et al. [12] also relied on external evidence to train an evidence-aware neural network model. Their system aggregated signals from the external evidence articles and the reliability of their sources. The evidence articles which are relevant to the claims, are collected using a search engine. Different from previous studies, we are interested in investigating the effectiveness of emotional signals extracted from the text of the claim in the credibility assessment task.

Emotions can play a critical role for addressing different tasks ranging from irony detection [5] to recommender systems [18]. There are many theories for emotions in psychology such as Ekman's theory, according to which, there are six core emotions: fear, anger, disgust, sadness, happiness and surprise [2]. Emotions can also play an important role in false information detection. Vosoughi et al. [16] investigated a large corpus of tweets and found that false rumours trigger fear, disgust and surprise in their replies, whereas the true rumours trigger joy, sadness, trust and anticipation. Several approaches have been proposed to annotate a text according to the emotion it expresses. In addition, several emotion lexicons were developed to facilitate emotion analysis. For example, EmoLex [9] contains terms that can be used to annotate documents as expressing an emotion such as anger and sadness. Other researchers [3, 4] focused on emotional reactions (love, joy, surprise, sadness, anger) that are triggered by news posts online and tried to predict the number of those reactions.

3 METHODOLOGY

We propose EmoCred that is an approach based on an LSTM neural network and which 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 $[w_1, w_2, ..., w_l]$ where $w_l \in \mathbb{R}^d$ and $\mathbb{R}^d$ is the d-dimensional word embedding of the lth word in the input claim c.

As a second input, we consider emotional signals that are extracted from the text of the claim. Section 3.1 describes 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:

$$d_e = relu(W_e(esv) + b_e)$$

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

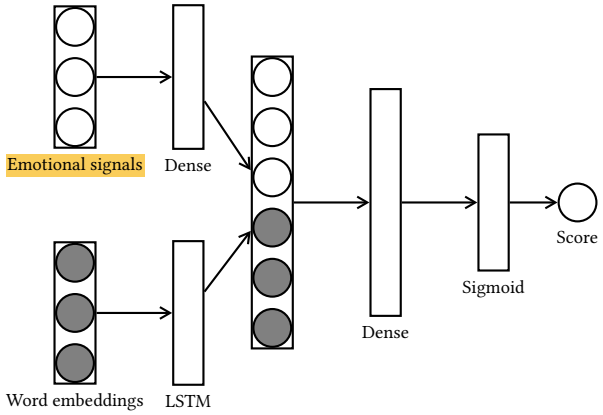

Figure 1: EmoCred neural network architecture for credibility assessment.

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:

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

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

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

$$S = sigmoid(d_m)$$

3.1 Emotional Signals

One important step of EmoCred is to extract the emotional signals from the claims. We explore three different approaches for calculating the emotional signals of the claims; (i) a lexicon based approach that is based on state-of-the-art emotional lexicons (*emoLexi*), (ii) an approach that calculates the emotional intensity of the claims (*emoInt*) and, (iii) a neural network approach that assigns an intensity level to the claims and which represents the number of emotional reactions that the claim can trigger to the readers (*emoReact*).

The first approach is straightforward and is based on emotional lexicons. This approach considers that 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 $\mathscr{E} = [t_{e,1}, t_{e,2}, ..., t_{e,L}]$ that convey a specific emotion e. Then the *emoLexi* approach considers the number of the emotional words that appear in the text of the claim. This process can be described as:

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

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:

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

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

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 if a claim c regarding a specific emotion e is calculated as:

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

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

The *emoReact* approach aims to predict the intensity level of emotional reactions that will be triggered to the readers. This approach predicts for each claim the probability to trigger any of the five intensity levels (very low, low, average, high, very high) for each of the five reactions love, joy, surprise, sadness and anger. The *emoReact* approach is based on a standard bidirectional LSTM network with attention. The neural network is trained with pre-trained word embeddings.

4 EXPERIMENTS

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

4.1 Datasets

For the evaluation we use data from PolitiFact[2] 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. We have decided not to merge the two datasets so we can use the pre-defined train, test and development sets.

The first dataset[3] (Politifact-1) was presented by Popat et al. [12]. The dataset contains 3,568 claims of which 1,867 are credible and 1,701 are not credible claims. The dataset also contains snippets from 29,556 articles that correspond to a total of 336 article sources. However, we do not use these snippets in our study.

The second dataset (Politifact-2) was presented by Rashkin et al. [14]. 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. Following previous studies [12, 14] we combine true, mostly true and half true labels into one class label (i.e., true) and the rest as false and, therefore, we treat the problem as a binary classification.

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 [9]: EmoLex contains the associations of 14,181 words with eight emotions (i.e., anger, anticipation, disgust, fear, joy, sadness, surprise, trust)
- SentiSense [1]: This lexicon is a concept based affective lexicon and contains associations between WordNet concepts and emotional categories
- EmoSenticNet [13]: A lexical resource that contains emotional labels for SenticNet concepts

For the *emoInt* approach, we use the NRC Affect Intensity Lexicon [8]. 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. The dataset is the same used by Giachanou et al. [4] 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. We train a standard bidirectional LSTM model with attention on the Facebook posts and then this model is used to predict the intensity of the emotional reactions in the claims. More specifically, for each of the claim, we predict the probability to trigger any of the five different intensities (very low, low, average, high, very high) of five different emotional reactions (love, joy, sadness, surprise, anger). We use the pre-trained GloVe Wikipedia 6B word embeddings [10] to initialise 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 rest 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 pre-trained GloVe Wikipedia 6B word embeddings [10] to initialise our word embeddings.

We use Keras with a Tensorflow backend to implement our system. All the neural networks models are trained with Adam optimizer with a learning rate of 0.002 to minimize binary cross-entropy loss. We use L2 regularizers with the fully connected layer. Also, we set dropout to 0.5. The hyper-parameters are trained on the development sets.

We report accuracy and macro F1-metric for the evaluation of the models. We compare the performance of EmoCred with that of LSTM-text approach proposed by Rashkin et al. [14]. This approach is based on an LSTM network trained using only word embeddings from the text of the claims, initialised using the pre-trained word embeddings. Finally, we use the McNemar test to measure the statistical difference.

5 RESULTS

Table 1 summarises the results of our experiments. From the results, we observe that EmoCred outperforms the LSTM baseline by a large margin. In all the cases, we observe that incorporating emotional signals into LSTM significantly improves the performance in the credibility assessment task. Regarding the Politifact-1 dataset, the best performance is achieved with the emotional reactions. In particular, EmoCred-emoReact manages to significantly outperform LSTM-text by 12.39% in terms of F1-score. This is a very interesting result because the *emoReact* is trained on different data which were crawled from Facebook. Even the model was trained on data from a different domain, it seems that still the emotional features are very helpful for the credibility assessment task. In case of *emoInt* and

[2]https://www.politifact.com/
[3]https://www.mpi-inf.mpg.de/dl-cred-analysis/

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%.

Similarly, EmoCred outperforms LSTM-text baseline on Politifact-2. In this case, the best performance is obtained with the *emoLexi* approach that significantly outperforms the baseline by 6.5%.

Table 1: Performance results of EmoCred approach when using different approaches for generating the emotional signals. A star(∗) indicates statistically significant improvement over the LSTM-text approach.

Dataset	Method	Accuracy	F1-score
Politifact-1	LSTM-text	0.551	0.549
	EmoCred-emoLexi	0.608	0.602∗
	EmoCred-emoInt	0.604	0.602∗
	EmoCred-emoReact	0.617	0.617∗
Politifact-2	LSTM-text	0.597	0.567
	EmoCred-emoLexi	0.621	0.606∗
	EmoCred-emoInt	0.628	0.586
	EmoCred-emoReact	0.619	0.601∗

Finally, we conduct a further analysis on the emotional signals to understand which emotions contributed more to the performance. For this reason, we run the EmoCred-emoLexi on the PolitiFact-1 dataset using groups of emotions instead of all the emotions. To be more specific, 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. [16] who found that false rumours trigger fear, disgust and surprise whereas true rumours trigger joy, sadness, trust and anticipation.

Table 2: Performance results of EmoCred-emoLexi approach when using different groups of emotional signals. *Group1* includes fear, disgust and surprise, whereas *group2* joy, sadness, trust and anticipation. A star(∗) indicates statistically significant improvement over the LSTM-text approach.

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

Table 2 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 improves the baseline in contrast to EmoCred-emoLexi (group1) that also improves the baseline but not significantly.

6 CONCLUSIONS

In this work, we propose the *EmoCred* approach, an LSTM-based model that combines information from the claims' text with emotional signals for credibility assessment. We study three different methodologies for generating the emotional signals from the text of the claim; lexicon-based emotional analysis, emotional intensity and a neural network for generating the intensity level of emotional reactions. Experiments on real world datasets show the effectiveness of emotional signals for the task of credibility assessment.

In the future, we plan to extend our model by using emotional features that are extracted from different sources such as external articles that are relevant to the claim. Finally, we plan to explore the effectiveness of ensemble neural networks on the credibility assessment task.

Acknowledgments.

The first author is supported by the SNSF Early Postdoc Mobility grant P2TIP2_181441 under the project *Early Fake News Detection on Social Media*, Switzerland.

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] J. Carrillo de Albornoz, L. Plaza, and P. Gervás. Sentisense: An easily scalable concept-based affective lexicon for sentiment analysis. In *LREC '12*, pages 3562–3567, 2012.
[2] P. Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3/4):169–200, 1992.
[3] A. Giachanou, P. Rosso, I. Mele, and F. Crestani. Early commenting features for emotional reactions prediction. In *SPIRE '18*, pages 168–182, 2018.
[4] A. Giachanou, P. Rosso, I. Mele, and F. Crestani. Emotional influence prediction of news posts. In *ICWSM '18*, pages 592–595, 2018.
[5] D. I. Hernańdez-Farías, V. Patti, and P. Rosso. Irony detection in twitter: The role of affective content. *ACM Transactions on Internet Technology*, 16(3):19:1–19:24, 2016.
[6] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
[7] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha. Detecting rumors from microblogs with recurrent neural networks. In *IJCAI'16*, pages 3818–3824, 2016.
[8] S. M. Mohammad. Word affect intensities. In *LREC '18*, pages 174–183, 2018.
[9] S. M. Mohammad and P. D. Turney. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465, 2013.
[10] J. Pennington, R. Socher, and C. Manning. Glove: Global vectors for word representation. In *EMNLP '14*, pages 1532–1543, 2014.
[11] K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum. Where the truth lies: Explaining the credibility of emerging claims on the web and social media. In *WWW '17 Companion*, pages 1003–1012, 2017.
[12] K. Popat, S. Mukherjee, A. Yates, and G. Weikum. Declare: Debunking fake news and false claims using evidence-aware deep learning. In *EMNLP '18*, pages 22–32, 2018.
[13] S. Poria, A. Gelbukh, A. Hussain, N. Howard, D. Das, and S. Bandyopadhyay. Enhanced senticnet with affective labels for concept-based opinion mining. *IEEE Intelligent Systems*, 28(2):31–38, 2013.
[14] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *EMNLP '17*, pages 2931–2937, 2017.
[15] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54, 2010.
[16] S. Vosoughi, D. Roy, and S. Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.
[17] W. Y. Wang. "liar, liar pants on fire": A new benchmark dataset for fake news detection. In *ACL (Volume 2: Short Papers)*, pages 422–426, 2017.
[18] Y. Zheng, B. Mobasher, and R. Burke. *Emotions in Context-Aware Recommender Systems*, pages 311–326. 2016.