Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

Emotional Fingerprints: A Framework for Identifying Fake News

Pavo Stanić
Monash University, Australia and University of Saskatchewan, Canada
pavo.stanic@monash.edu
Marjan Nikolić
Swiss Federal Laboratories for Materials Science and Technology.
marjan.nikolic@empa.ch

Abstract

Social media has become one of the main channels of information for human beings due to the immediacy and social interactivity they offer, allowing, in some cases, the publication of whatever each user considers relevant. This has led to the generation of false news, or fake news, publications that only seek to generate uncertainty, misinformation, or bias readers' opinions. It has been shown that humans are not able to fully identify whether an article is factual or fake news. Because of this, models have emerged that seek to characterize and identify articles based on data mining and machine learning. This article proposes a three-layer framework, the main objective of which is to characterize the emotions present in fake news and to serve as a tool for associating the emotional state and the most likely intention of the person publishing a fake news story.

Keywords: Fake news; emotions; characterization of fake news

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

Introduction

The use of social networks and the presence of Fake News have generated a negative impact on society [1], generating uncertainty and misinformation, in addition to this it has been shown that human beings are not able to identify with high certainty whether an article is really a fact or Fake News, due to this, the problem of misinformation associated with the generation and spread of Fake News has become relevant at the level of states and countries that are being affected by this phenomenon of human origin but aggravated by technologies that facilitate its increase.

The state of the art accounts for works that seek to characterize Fake News in various ways: textual characteristics [2], verbal tenses [3], cognitive complexity [4], categorization of positions [5], emotional polarity [6], [7], among others.

The contribution of this article is based on the development of a framework for identifying emotions present in social media posts. Applied to the field of fake news, this framework allows us to characterize not only the emotions present, but also how these characteristics vary compared to non-fake posts.

In this sense, from the point of view of the purpose of the research, this work seeks to answer the following research questions:

RQ1: *How can emotions associated with certain words present in the text be incorporated into a text that represents general news?*

RQ2: *How can we quantify the degree of emotionality of each type of news, ie emotionality present in Fake News versus that present in Non-Fake News?*

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

RQ3: *How can we contrast the variations in emotions expressed in a set of news items from one domain versus those expressed in a different domain?*

The rest of the article is organized as follows: first, the general architecture of the proposed framework is described, as well as the modular layers that comprise it.

The results obtained from applying this framework to three news sets from different domains are presented: Politics, Celebrities, and COVID-19. The results are then discussed in relation to the previously formulated research questions. Finally, the conclusions of the work developed are presented, highlighting its main limitations and identifying possible lines of future research.

Proposed Framework

The proposed Framework *(* Figure 1 *)* seeks to identify emotions in a set of publications, for which it is proposed that its structure correspond to three layers or modules: Layer 1: Emotionalization, Layer 2: Quantification, and Layer 3: Characterization.

Each of these layers works based on the words used, the type of publication, and the domain to which they belong.

Emotionalization

The first layer of the proposed framework requires two datasets, NRC-Emotion Lexicon, a dataset that incorporates words, emotions, and the intensity with which they relate to each other [8] . The second dataset must be related to the publications that will be analyzed based on emotions. A prior phase of text data preparation is assumed for this dataset, except for the application of Lemmatization and Stemming, whose application is not appropriate in this case.

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

In 2019, Anoop, Deepak, and Lajish presented a way to incorporate emotions into health-related Fake News in order to improve their classification models [9]. The implemented method seeks to compare each word of a post with the NRC-Emotion Lexicon. Emotionalization can be represented by equation (1), where *W* is a word in a post, *E is* the emotion with which it could be related, and *I* corresponds to the intensity of the relationship. Previously, it is required to establish a threshold (τ) so that, if the intensity of the emotion *E* with respect to the word *W* turns out to be greater than or equal to τ , *then the word W* is added to the corpus followed by the emotion E, otherwise only the word *W is kept*, without any added emotion.

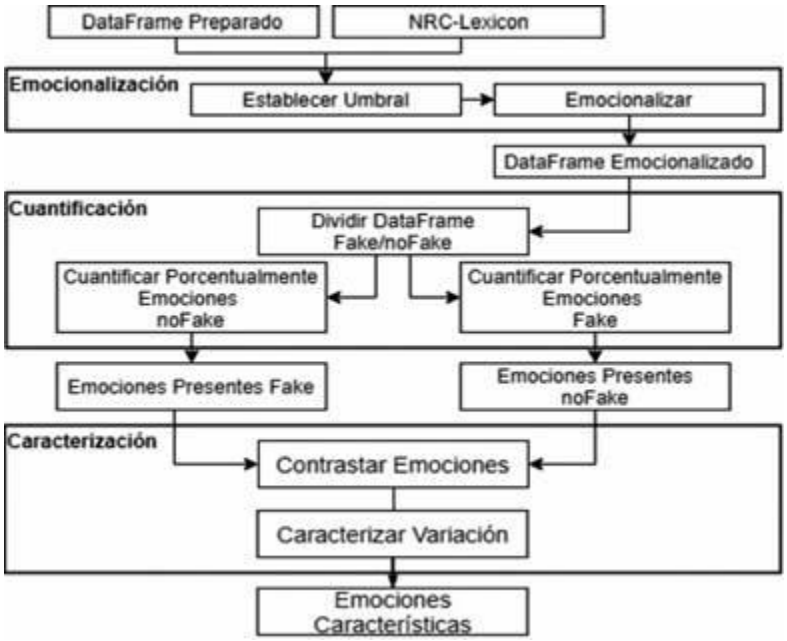

Figure 1 General architecture of the proposed framework.

$$f_{(W,E,I)} = \begin{cases} W+^{""}+E & I \geq \tau \\ W & I < \tau \end{cases} \quad (1)$$

Quantification

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

The quantification layer starts with the emotionalized DataFrame generated by the Emotionalization layer, which is divided into two new sets based on the type of publication; Fake and noFake. After that, each DataFrame is subjected to a quantification process, which can be expressed according to equation (2), where $E_{(e,i)}$ represents the frequency of an emotion *(e)* in a type of publication (*i*), and $W_{(i)}$ the number of words per type of publication (*i*). As an output of this layer, a matrix with the percentage frequency of the emotions present by type of publication is calculated, this for each different domain.

$$P_{(e,i)} = \frac{E_{(e,i)} * 100}{W(i)} \quad (2)$$

Characterization

The characterization layer consists of two steps: 1) Emotion contrast and Variation characterization. The emotion contrast can be quantified by means of the calculation defined in equation (3), where the variation of an emotion ($V_{(e)}$) is defined as the difference between the percentage quantification of the emotion *e* for nonFake posts ($P_{(e',noFake')}$) and the percentage quantification of the same emotion *e* for Fake posts ($P_{(e',Fake')}$).

$$V_{(e)} = P_{(e,'noFake')} - P_{(e,'Fake')} \quad (3)$$

Furthermore, variation characterization allows for the graphical identification of detected variations. To do this, a variation matrix is calculated using the calculation defined in equation (4), where:

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

$$C_{(e,\mu)} = \begin{cases} \downarrow & V_{(e)} > \mu \\ \uparrow & V_{(e)} < -\mu \quad |\mu > 0 \\ = & \mu \geq V_{(e)} \geq -\mu \end{cases} \tag{4}$$

If $V_{(e)} > \mu$, it means that an emotion *e* is presented less in Fake posts.

If $V(_e) < -\mu$, it translates as that an emotion *e* is presented more in Fake posts.

If $\mu \geq V(_e) \geq -\mu$, it is interpreted as that an emotion in is presented in equal measure in Fake and nonFake posts.

Application of the Framework

In order to validate the use of the presented framework, three datasets are collected: PolitiFact, GossipCop and CoronaFake, where the first dataset is related to the domain of Politics, the second to Celebrities and the third to Covid-19. The three sets address both Fake and non-Fake publications, these being collected from *X* (formerly Twitter). It should be noted that the three datasets have been previously preprocessed, removing URLs, converting everything to lowercase, eliminating StopWords and removing duplicates. The distribution of publications after data processing can be seen in Table 1 .

Table 1 Distribution of publications by Dataset.

Dataset	Fake	No Fake
PolitiFact	130.555	448.397
GossipCop	460.187	445.332
CoronaFake	575	584

Table 2 shows the result of the quantification ($P_{(e,\ i)}$), showing how frequent an emotion is relative to the total number of words and the type of publication. The type of publication in a given domain where an emotion is most frequent is shown

Technology Journal of Management , Accounting and Economics (TECH)
ISSN: 2311-3995

Vol. 13 No. 2 (2025)

in bold. The type of publication in the domain where the emotion is most frequent is underlined.

Table 2 Quantification results by domain and type of publication.

Table 3 shows the application of the characterization layer, evidencing the change in emotions present between Fake and non-Fake news from each of the three domains explored in this work, that is, Politics, Celebrities and Covid-19.

Table 3 Results of the characterization by domain.

Emoción	Política	Celebridades	Covid-19
Anger	↑	↑	↑
Anticipation	=	↓	↓
Disgust	↑	↓	↓
Fear	↑	↓	↑
Joy	↓	↓	↑
Sadness	↑	↓	↑
Surprise	↑	↓	↑

DISCUSSION OF RESULTS

In terms of the results presented in the application of the proposed framework to three sets of news from three different domains, and based on the research questions formulated to guide the experimental work of the research, the following can be seen:

RQ1: ¿Cómo se puede incorporar a un texto que representa una noticia en general, emociones que se asocian a ciertas palabras presentes en el texto?

The proposed framework addresses the incorporation or aggregation of emotions to the original text by using the NRC-Emotion Lexicon, to search for which emotions

are strongly associated with a word, for which an intensity threshold is defined so that the word-emotion relationship can be considered significant. In that case, the emotion with which it is most strongly connected and that exceeds the defined threshold is added to the original word in the original text. This is done for both fake news and non-fake news. Experiments are made with different intensity thresholds (e.g., 0.2, 0.4, 0.6, 0.8), and empirically, it is found that the intensity threshold of 0.6 is appropriate for incorporating significant emotions into the words that make up the news. Thus, the results presented in Tables 2 and 3 are determined based on an intensity threshold of 0.6, in the relationship between words in each news item and the emotions listed in Tables 2 and 3 .

RQ2: ¿Cómo se puede cuantificar el grado de emocionalidad de cada tipo de noticias, i.e. emocionalidad presente en Fake News versus las presentes en No Fake News?

The percentage in which a certain emotion e is present in a type of news i is defined , for a given domain, as formulated in equation (2). When this was applied to the three different domains, the results summarized in Table 2 were obtained . Here it is identified that, for the Politics domain, in the Non-Fake News the emotion Joy is more predominant than the other emotions, but in the Fake News the emotions Fear and Sadness predominate. In relation to the Celebrities domain, in the Non-Fake News the emotions Joy and then Anticipation are more predominant, but in the Fake News Fear stands out as differentiating with respect to the Non-Fake News in this domain. In relation to the Covid-19 domain, the emotion Disgust stands out for having a higher percentage in No-Fake News than

its corresponding value in Fake News, but in the Fake News the emotions Fear and Sadness stand out in this domain.

RQ3: ¿Cómo se puede contrastar las variaciones de las emociones expresadas en un conjunto de noticias correspondientes a un dominio versus las expresadas en otro dominio distinto?

The calculation of the percentage variation between emotions expressed in Fake News versus Non-Fake News in a particular domain is defined, with which a matrix of this percentage variation is constructed in terms of whether an emotion grows (↑), decreases (↓), or remains in the range [-μ, +μ], defined by the parameter μ>0. In terms of the 3 domains, the emotions Fear, Sadness, and Surprise experience growth in the Politics and Covid-19 domains, unlike the Celebrities domain, which experiences a decrease in Fake news with respect to Non-Fake news. Therefore, in terms of these 3 emotions, the Politics and Covid-19 domains exhibit similarity between them, but are dissimilar with respect to the Celebrities domain (Table 3). In turn, Celebrities is similar in the Disgust emotion with the Covid-19 domain, and with the Joy emotion with the Politics domain.

Conclusion

This article proposes a framework for emotion identification in Fake News from a given domain and presents its application in three different domains. Primarily, this work contributes to the state of the art with a methodology for detecting emotions in different domains, recognizing specific emotions rather than polarities (positive, negative, neutral). This framework could bring benefits such as improved Fake News classification rates by incorporating key emotions as a discriminating

element, early Fake News detection by identifying the high predominance of an emotion usually associated with Fake News in the context of a given domain, identifying the underlying emotional state associated with a certain probable intention of the person publishing Fake News, or in other areas such as the generation of advertising campaigns, speeches, or any media that requires addressing specific emotions.

As future work, we hope to incorporate tools into the framework that allow us to identify not only emotions but also the intentions of the poster, seeking to establish quantifiable relationships between the two. This will require defining new metrics to measure the intensity of the relationship between emotion and intention in fake news.

References

[1] L. Rodríguez-Fernández, "Disinformation and organizational communication: a study on the impact of fake news," *Latin American Journal of Communication* , no. 74, pp. 1714-1728, doi: 10.4185/RLCS-2019-1406.

[2] B. Horne and S. Adali, "This Just In: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news," *in Proceedings of the International AAAI Conference on Web and Social Media(ICWSM)*, vol. 11, no. 1, Montreal, Quebec, Canada, 2017, pp. 759-766, doi: 10.1609/icwsm.v11i1.14976.

[3] V. Perez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," 2017, arXiv: 1708.07104.

[4] N. Nguyen, G. Yan, M. Thai, and S. Eidenbenz, "Containment of misinformation spread in online social networks," in *Proceedings of the 4th Annual ACM Web Science Conference, Association for Computing Machinery*, Evanston, Illinois, USA, 2012, pp. 213-222, doi: 10.1145/2380718.2380746.

[5] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and Resolution of Rumours in Social Media: A Survey," *ACM Computer Surveys*, vol. 51, no. 2, Art. no. 32, 2018, doi: 10.1145/3161603.

[6] F. Qian, Ch. Gong, K. Sharma, and Y. Liu, "Neural user response generator: Fake News detection with collective user intelligence," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, (IJCAI), vol. 18, 2018, pp. 3834-3840, doi: 10.24963/ijcai.2018/533.

[7] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context, and spatio temporal information for studying fake news on social media," *Big Data*, vol. 8, no. 3, pp. 171-188, 2020, doi: 10.1089/big.2020.0062.

[8] S. Mohammad and P. Turney, "Crowdsourcing a Word-Emotion Association Lexicon," *Computational Intelligence*, vol. 29, no. 3, pp. 436-465, 2013, doi: 10.1111/j.1467-8640.2012.00460.x

[9] K. Anoop, P. Deepak, and V.L. Lajish, "Emotion cognizance improves health fake news identification," 2020, arXiv: 1906.10365.