4th International Conference on Advanced Research Methods and Analytics (CARMA2022)
Universitat Politècnica de València, València, 2022
DOI: http://dx.doi.org/10.4995/CARMA2022.2022.15081

Machine Learning and MADIT methodology for the fake news identification: the persuasion index

Luisa Orrù[1], Christian Moro[1], Marco Cuccarini[2], Monia Paita[1], Marta Silvia Dalla Riva[1], Davide Bassi[1], Giovanni Da San Martino[2], Nicolò Navarin[2], Gian Piero Turchi[1]
[1]Philosophy, Sociology, Pedagogy and Applied Psychology, University of Padova, Italy,
[2]Mathematics Department, University of Padova, Italy-

Abstract

The phenomenon of fake news has grown concurrently with the rise of social networks that allow people to directly access news without the mediation of reliable sources. Recognizing news as fake is a difficult task for humans, and even tougher for a machine. This proposal aims to redesign the problem: from a check of truthfulness of news content, to the analysis of texts' persuasion level. That is how information is introduced to the reader, assuming that fake news is aimed at persuading towards the reality of sense they intend to convey. M.A.D.I.T. methodology has been chosen. It is useful to describe how texts are built, overcoming the content/structure analysis level and stressing the study of Discursive Repertories: discursive modalities of reality of sense building, classified into real and fake news categories thanks to the Machine learning application. For the dataset building 7,387 news have been analysed. The results highlight different profiles of text building between the two groups: the different and typical discursive repertories allow to validate the methodological approach as a good predictor of the persuasion level of texts, not only of news, but also of information in domains such as the economic financial one (e.g. GameStop event).

Keywords: *Fake news; Persuasion index; MADIT methodology; Machine learning; Dialogic analysis; Discursive configuration.*

1. Introduction

In recent years, the phenomenon of fake news showed its critical effects (Tandoc, et al., 2018; Tagliabue, et al., 2020). In light of this, the scientific community has worked to provide tools to help citizens identify fake news. The scientific efforts in contrasting fake news outcomes have been undertaken in two “main directions” (Lazer et al., 2018): empowering individuals in evaluating the fake news they encounter, and implementing structural changes on online platforms and algorithms, to prevent exposure of individuals. Referring to the first category of interventions, different studies focused on finding personal characteristics and cognitive processes that play a role in dealing with fake news (Pennycook, Cannon & Rand, 2018; Pennycook & Rand, 2020). About the second category of interventions, Artificial Intelligence and Natural Language Processing have become increasingly important tools to help citizens when dealing with fake news (Oshikawa, et al, 2020). Currently, several different computerized methods are available for detecting fake news (for a review, see Zhou & Zafarani, 2020; Oshikawa, et al., 2020). The plurality of available methods can be traced back to the lack of a specific and shared definition of "fake news", which currently assumes different characteristics depending on the author or the research considered (Tandoc, et al., 2018; Andersen & Søe, 2020). At the same time, Zhou e Zafarani (2020) offer valuable support, organizing all the constructs “similar to fake news (es. satire, clickbait, etc) in function of their intention and their truthfulness. However, some questions are left open: how is the truth of a certain content decided, when the news cannot find factual correspondences (Andersen & Søe, 2020)? How much of the analyzed content is false, and how to consider the news when it’s partially false (Oshikawa, et al., 2020)?

The mere truthfulness of a piece of content does not offer any insight into how it is conveyed by the text: the presented work attempts to abandon the dichotomy "fake-news/real-news" in favor of the construct of *persuasion*. In doing so, we do not intend to replace content truthfulness analysis with persuasion analysis, but rather to shift the focus of the investigation: from the reader who assumes this information to be fake or true, to the degree to which the modalities used to convey the text lead the reader to assume the same narrative position, i.e. the same modalities used in and by the text itself. It is possible to anticipate how fake news conveyed through highly persuasive modes can lead readers, and more broadly the community, towards more high-risk interactive settings, e.g. of social fragmentation. This perspective is in continuity with Barron-Cedeno et al. (2019) contribution on propaganda (Da San Martino et al., 2020): the authors move beyond the fake/real news distinction, in favor of an approach that provides an index related to how much the news tries to influence the reader's opinion through automated analysis of the structure and content of the text: a persuasion index (Miller & Levine, 2019; Festinger 2001; Grandberg, 1982; Cacioppo, Cacioppo, & Petty, 2018; Druckman, 2021). O'Keefe’s perspective on persuasion (2016), integrated with the analysis tools made available by the

NLP, enables the extraction of the characteristics of the arguments that allow us to effectively change the interlocutor's "position" on a given argument (“persuasion techniques”'; Li, et al., 2020; Hunter et al., 2019). Following these recent works, the analysis of the ways in which language is used enables the identification of typical modes of persuasive messages. The focus of our work is the same ways of using language: the persuasiveness of a text, which is in fact associated with the rhetorical-argumentary architecture of the text itself (how it is discursively structured) and the amount of critical reading competence and attention needed to consume it. As a theoretical reference to solve this problem, we have chosen Dialogical Science (Pinto et al., 2022; Turchi, et al., 2021), which, through the formalization of Natural Language, has given value to the description of Natural Language’s use, formalizing the rules of its use. The formalization of language has given value to saying, defining and grouping it, and then measuring it (Turchi and Orrù, 2014). Following the approach of Dialogical Science, we defined persuasiveness as “the possibility of a text to make the reader use discursive modes and contents similar to those of the text itself”[1]. This possibility of the text is therefore independent of the truthfulness of the contents; however, we anticipate that fake news disseminates texts that are more prone to this possibility. We then defined a novel index of persuasiveness of a text derived from the Discursive Repertories (DRs; Turchi et al., 2021; Turchi and Orrù, 2014), namely specific language use modalities as defined by M.A.D.I.T methodology (Turchi et al., 2021). There are 24 DRs, each identifying one of the possible language use modalities that can be traced in a text. Based on the specific characteristics and properties of each DR, we hypothesized that 12 DRs can be understood as the elements that constitute the DNA of a persuasive news, since facts can be manipulated through language in order to convince the reader about some version of them, creating also what we called fake news (Iswara and Bisena, 2020). This exploratory experimentation wanted to observe whether the DRs that, theoretically, should characterize persuasive texts, characterize the texts of fake news more than the one of real news. We anticipate that even real news texts could employ DRs related to the construct of persuasion, but we consider that fake news and persuasion index are related, certainly it will be the object for a future work of deepening.

2. Methods

In order to pursue the goal of this exploratory experimentation, building a dataset containing both fake news and real news was necessary. We define real all the news taken by authoritative Italians newspaper and fake all the news taken from list of blogs and web sites

[1] Rephrasing of the technical definition of Persuasion: "*Possibility of a text to generate a configuration of sense tending in turn to occupy - without ever overlapping - the same discursive space of the original text*”.

that spread medical, scientific and political misinformation. The list was provided by the fact checking web site: Bufale.net. To obtain the texts of these articles automatically, we used a package of the python library which, after providing the link of the news' container site, returns the title, the text and other information of the news itself. We found an imbalance in the topics covered: fake news usually focuses on a certain set of topics, ignoring others. To overcome this issue, we eliminated from real news the topics that are rarely covered in fake ones, for example sports. The dataset is composed of 2776 real news and 4611 fake news. We built a corpus of human annotated texts and devised a machine learning approach to identify the DRs. Naive Bayes was initially employed, but its low precision rate (0.35) and recall rate (0.37) led us to use BERT. We created a model which manages to divide the inserted text into excerpts. Subsequently, BERT classifies these excerpts according to the 24 possible categories (as the DRs). The model that is closest to human performance has a precision level of 0.47 (recall=0.43; f1-score=0.43), which can be improved considering that human roles trained with a basic training have precision=0.65; recall=0.63; f1-score=0.63. We use a model previously trained for the identification of DRs, that model is a BERT. For identification of DRs, we used a dataset of 14567 excerpts, each of them belongs to one of the 24 possible categories (as the DRs). We have defined different possible models and using the cross-validation we chose the best one in terms of performance. In our case was that one is described in the Table 1. Chose the best model, we split the dataset in train (75%) and test (25%), we train it and we get the result.

Table 1. BERT model structure

Model structure	Pretrained Weights	Batch Size	Learning Rate	Freeze to
The structure of model that produce the pertrained weights.	bert-base-italian-xxl-uncased	16	1e -05	1

AdamW Eps	Max Epochs	Patience Epochs	Embed Dim	Activation Function
0.0001	200	20	768	ReLU

For the construction of the dataset with the DRs distribution, we implement a model for text division in extracts, we used the function sent_tokenise of the package NLTK. This tokenizer divides a text into a list of sentences by using an unsupervised algorithm. Following the construction of the dataset, we used the model to analyze the texts of fake and real news, to detect how the chosen model divides the various excerpts and names them, going into detail of the distribution of the DRs for fake and real news. Having for each text, real or fake, the distribution of DRs, we used a model to predict whether a text is a real news or a fake news;

this, to provide further support for the theoretical link between persuasion and biased texts, based on the distribution. We then implemented a Random Forest model, which allows us to define the importance that each repertory has had in the classification. Either in this case we spit the data in train (75%) and test (25%) for searching the best hyper parameters, we used RandomisezSearchCV of the package Sklearn. Hyper-parameters are optimized by cross validated search over parameter settings.

3. Results

Figure 1 shows the distributions of RDs according to the text categorization in fake and real news. Generally, a strong presence of characteristic repertories is observable for both the text codifications.

Table 2 addresses the Figure 1's distributions, and reports the percentage occurrences of the 12 DRs in fake and real texts. We use the distribution to characterize a persuasive text. The assumption of our research was that we would be able to trace an increased occurrence of persuasion-indicating DRs regarding the texts of fake news, more than the real news ones. The results confirm this hypothesis for seven of the twelve indicated DRs ("Anticipation", "Cause of Action", "Declaration of Aims", "Prediction", "Justification", "Certify Reality" and "Evaluation").

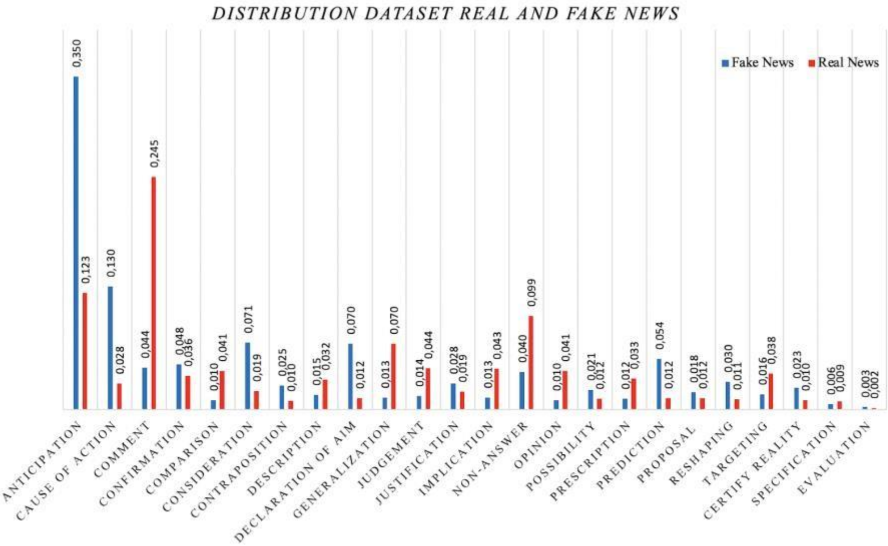

Figure 1. Distribution Dataset of real and fake news

Table 2. Distribution of indicators of persuasions between fake and real news

DR	Fake	Real	RF	DR	Fake	Real	RF
Anticipation*	0.35	0.123	0.00937	Targeting	0.016	0.038	0.01921
Cause of Action*	0.13	0.028	0.03254	Judgment	0.014	0.044	0.03176
Declaration of Aim*	0.07	0.012	0.02566	Implication	0.013	0.044	0.03111
Prediction*	0.054	0.012	0.05226	Prescription	0.012	0.033	0.03958
Justification*	0.028	0.019	0.02134	Opinion	0.01	0.041	0.02325
Certify Reality*	0.023	0.01	0.03547	Evaluation*	0.003	0.002	0.02545

When looking at the results, it is also necessary to consider the imbalance between fake and real news in the dataset, the level of accuracy of the model with respect to the naming task (0.47) and, in general, the errors brought behind in the various phases of the program, both by the model that classifies the DRs but also from the one that divides the text. The results in the prediction of a text as real or fake news are with the Random Forest' model: Precision=0.76; Recall=0.76, F1-score=0.76; which can be considered very good results due to the difficulty of the task and the margin of error of the model that predicts the DRs of each text.

4. Conclusions

In line with the work of Barron-Cedeno and colleagues (2019), a perspective was adopted that attempted to overcome the issues of the traditional 'fake-news/real-news' distinction. Therefore, a persuasion index was constructed to provide the reader with elements to evaluate the bias of a certain piece of news, based on the specific ways in which language is used in a text. The analysis of language use was carried out according to Dialogic Science (Turchi and Orrù, 2014), and implemented through an automatic process, based on the BERT transformer. The results obtained from the experiment generally support what has been argued in theory. Specifically, we observed a greater occurrence of DRs - which theoretically should generate persuasion in the reader - in fake news texts, and a greater occurrence of "less persuasive" DRs in real news texts, as was also shown by the results of the Random Forest model. Therefore, the elements of the 'DNA' of persuasive news, identified through the Dialogic Science, have been matched by experimental data. However, this kind of distribution, in support of theory, has not been found in all repertories that are supposed to promote a process of persuasion. This can be traced back to several factors: the imbalance of the dataset, the error made by the model in processing the text in excerpts, as well as in the naming process. Thus, the findings highlight certain aspects of the model used and the need of the automated textual analysis to be specified and refined in the future, to increase the accuracy of the automated analysis. A further future work perspective concerns the construction of the dataset. As a matter of fact, the need for pre-processing work on the fake news dataset emerged: on the one hand, by maintaining a similar distribution between the fake news part and the real part; on the other hand, by refining the process of "cleaning" the text from parts that do not relate to the article in question, but are mistakenly downloaded from the Python library. Lastly, the index was applied in the economic-financial domain, to about 100 posts (on Reddit and Twitter) regarding the GameStop case (January 2021). A multitude of small investors, gathered on the Reddit page r/WallStreetBets, bought the company's shares in a mass movement that led them to rise, in less than a month, from $17.25 to $348, controversially setting up the entire financial market. It emerged that about 70% of the DRs used were indicators of persuasiveness: the posts addressed users in a uniform and coordinated way towards the goal of "opposing" Wall Street and big investors. This scenario could have been anticipated, observed and measured thanks to the persuasion index. The same index could be applied in the detection of war propaganda and in politics, since public convincing becomes the main aim of the professional roles involved, in order to gather consensus in a strategic way, avoiding risks of losing support (Durante and Zhuravskaya, 2018).

References

Andersen, J., & Søe, S.O. (2020). Communicative actions we live by: The problem with factchecking, tagging or flagging fake news – the case of Facebook. *European Journal of Communication,* 35(2), 126-139.

Barron-Cedeno, A., Jaradat, I., Da San Martino, G., & Nakov, P. (2019). Proppy: Organizing the news based on their propagandistic content. *Information Processing and Management,* 56, 1849-1864.

Cacioppo, J.T., Cacioppo, S., & Petty, R.E. (2018). The neuroscience of persuasion: A review with an emphasis on issues and opportunities. *Social Neuroscience,* 13(2), 129-172.

Da San Martino, G., Shaar, S., Zhang, Y., Yu, S., Barrón-Cedeno, A., & Nakov, P. (2020, July). Prta: A system to support the analysis of propaganda techniques in the news. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 287-293).

Druckman, J.N. (2021). A Framework for the Study of Persuasion. *Annual Review of Political Science,* 25.

Durante, R. & Zhuravskaya, E. (2018). Attack When the World Is Not Watching? US News and the Israeli-Palestinian Conflict. *Journal of Political Economy,* 126(31), 1085-1133.

Festinger, L. (2001). *Teoria della Dissonanza Cognitiva.* Franco Angeli.

Grandberg, D. (1982). Social Judgment theory. *Annals of the International Communication Association,* 6(1), 304-329.

Hunter, A., Chalaguine, L., Czernuszenko, T., Hadoux, E., & Polberg, S. (2019). Towards computational persuasion via natural language argumentation dialogues. In *Joint German/Austrian Conference on Artificial Intelligence* (Künstliche Intelligenz) (pp. 1833). Springer, Cham.

Iswara, A.A., & Bisena, K.A. (2020). Manipulation And Persuasion Through Language Features In Fake News. *RETORIKA: Jurnal Ilmu Bahasa*, 6(1), 26-32.

Lazer, D.M., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., ... & Zittrain, J.L. (2018). The science of fake news. *Science*, 359(6380), 1094-1096.

Miller, M.D., & Levine, T.R. (2019). Persuasion. In D. W. Stacks, M. B. Salwen, & K. C. Eichhorn, *An Integrated Approach to Communication Theory and Research* (p. 261 - 277). New York: Routledge.

O'Keefe, D.J. (2016). *Persuasion. Theory and Research.* London: Sage.

Oshikawa, R., Qian, J., & Wang , W.Y. (2020). A Survey on Natural Language Processing for Fake News Detection. *LREC,* 6086-6093.

Pennycook, G., Cannon, T.D., & Rand, D.G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of experimental psychology: general*, 147(12), 1865.

Pennycook, G., & Rand, D.G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of personality*, 88(2), 185-200.

Pinto, E., Alfieri, R., Orrù, L., Dalla Riva, M.S., & Turchi, G.P. (2022). Forward to a methodological proposal to support cancer patients: the Dialogic contribution for the precision care. *Medical Oncology,* 39(5), 75.

Tagliabue, F., Galassi, L., & Mariani, P. (2020). The "Pandemic" of Disinformation in COVID-19. *SN comprehensive clinical medicine,* 2(9), 1287-1289.

Tandoc, E.C., Lim, Z. W., & Ling, R. (2018). Defining “Fake News” A typology of scholarly definitions. *Digital journalism,* 6(2), 137-153.

Turchi, G.P., Dalla Riva, M.S., Ciloni, C., Moro, C., & Orrù, L. (2021). The Interactive Management of the SARS-CoV-2 Virus: The Social Cohesion Index, a Methodological Operational Proposal. *Frontiers in Psychology*, 12.

Turchi, G.P., & Orrù, L. (2014). *Metodologia per l'analisi dei dati informatizzati testuali: fondamenti di teoria della misura per la scienza dialogica*. Edises.

Zhou, X., & Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Survey,* 53(5), 40.