The Disinformation Battle: Linguistics and Artificial Intelligence Join to Beat it

La Batalla de la Desinformación: la Lingüística y la Inteligencia Artificial se Unen para Vencerla

Alba Bonet Jover
Department of Software and Computing Systems, University of Alicante, Spain
alba.bonet@dlsi.ua.es

Abstract: At present, the era of digitalisation has led to the era of disinformation by witnessing a peak of the spreading of fake news that are disseminated in order to get profit. Like everything, fake news has an Achilles' heel and it could be language. The linguistic structure as well as the expression of emotions through language could be key in detecting deception. The modelling of a language of deception typical of fake news and its later automation through machine learning would allow to take a further step towards the fight against disinformation.
Keywords: Natural Language Processing, Human Language Technologies, Fake News, Linguistic modelling, Machine Learning

Resumen: En la actualidad, la era de la digitalización ha dado paso a la era de la desinformación, presenciando así un auge en la viralización de noticias falsas que son difundidas con el fin de obtener un beneficio. Al igual que todo, las noticias falsas tienen un talón de Aquiles y este podría ser el lenguaje. La estructura lingüística utilizada así como la expresión de las emociones a través del lenguaje podrían ser clave en la detección de la mentira. La modelización de un lenguaje de la mentira propio de las noticias falsas y su posterior automatización mediante aprendizaje automático permitiría dar un paso más en la lucha contra la desinformación.
Palabras clave: Procesamiento del Lenguaje Natural, Tecnologías del Lenguaje Humano, Fake News, Modelización lingüística, Aprendizaje Automático

1 Justification

The term "fake news" is no longer an unfamiliar concept. Society is increasingly aware of the problem of the fast viralization of fake news in many areas of our daily life.

Lies have always existed, from those used to make political propaganda to win wars, such as the Goebbels' Nazi propaganda, to those used to win elections, as in the case of the Trump's victory. In fact, "the power of written texts to influence thought and opinion about someone or something has always existed in social life"(Nycyk, 2015). Lies used to obtain an economic goal have also circulated throughout history, such as those related to certain products sold as "miracle cures" and those that play with people's emotions in the health domain. In addition, there have always been lies that create ideological prejudices, such as the religious ones.

Fake news have always had different means of dissemination, but the development of new technologies and the arrival of Internet have created a breeding ground for the spreading of fake news. For that reason, this project expects to tackle this problem in the same environment it was created by using the Human Language Technologies.

Our proposal is based on the modelling of a language of deception by using both the manual and the automatic processes: the manual one would provide the automatic one the necessary information to learn the key features that could make the fake news detection possible.

2 Background and related projects

Nowadays, the term "fake news" is becoming increasingly popular. It was chosen as

Proceedings of the Doctoral Symposium on Natural Language Processing from the PLN.net network (PLNnet-DS-2020)

the Collins Word of the Year 2017 and it is defined by this dictionary as "false, often sensational, information disseminated under the guise of news reporting". Sukhodolov y Bychkova (2017) defined it as "a piece of news which is stylistically written as real news but which is completely or partially false". This last point is essential, as the concept "partially false" is the particularity that makes fake news difficult to detect and to deal with. According to Alba-Juez y Mackenzie (2019), "in most cases, fake news is not totally false, but rather a distorted version of something that really happened or a manipulated account of true facts". There are already studies that focus on this problem but, to the best of our knowledge, the structure and elements of the content of news are considered as a whole, so there is no system that detects the veracity of each part or essential element separately.

Other projects have analysed linguistic and emotional aspects that could be important when verifying the credibility of a news piece. Words are powerful in shaping people's beliefs and lexical features can help us to differentiate more reliable and less reliable digital news sources (Rashkin et al., 2017). By means of the language, fake news can play with people's emotions, which is very dangerous, since "emotion is inextricably linked to persuasion" and "persuasion can be used to heighten readers' sensitivity to a given issue, or on the contrary, to manipulate their emotions, stances and beliefs" (Alba-Juez y Mackenzie, 2019).

Society is becoming aware of the big problem of disinformation and more and more fact-checking agencies are fighting against lies and hoaxes. However, digital media is increasingly powerful and due to the fast spreading of fake news, the manual detection is an impossible challenge. For this reason, it is necessary to face the enemy in the same environment it was originated, in this case, in the digital world. Therefore, Artificial Intelligence must play a key role in the battle against disinformation, especially by applying the Human Language Technologies for the automatic detection of fake news. Some of the advances that have been made in NLP, specifically in linguistics and artificial intelligence domains, are approaches that combine linguistic cues and machine learning with network-based approaches (Conroy, Rubin, y Chen, 2015), methods using deep learning in discourse-level structure analysis (Uppal, Sachdeva, y Sharma, 2020), deep learning approaches for address stance detection (Borges, Martins, y Calado, 2019) and other tasks in fake news detection such as deception detection (by detecting the scarcity of information via linguistic cues); stance detection, controversy and polarization (by taking into account topics, impact and viralization of news), automated fact-checking, clickbait detection and approaches for measuring credibility (Saquete et al., 2020).

3 Hypothesis

3.1 Creation of a language model of deception based on the structure and content of news

Due to the fast dissemination of hoaxes and lies, it is important to fight against disinformation in order to provide society evidence-based information so people cannot make decisions that could harm their lives and those around them. Considering that a fake news contains both false and true information, the aim of this project is to create a language model based on the structure and the essential content elements of news, from which we will extract not only linguistic characteristics but also particularities of the context of the elements of fake news. Language is the skeleton of any news piece, it is its means of expression, so the hypothesis of this thesis is that through the linguistic, structural and content analysis of a fake news piece, elements that reveal its falseness could be detected.

Another important component in the creation of this language model is the emotional component. It is also reflected in language, in the way things are told. The hesitation, the inaccuracy or even the intention also present their own linguistic characteristics, either by the lexicon used, the grammatical mistakes, the use of undefined determinants or poorly constructed syntactic structures or the insistence on highlighting the semantic component of an idea.

3.2 Translation of news

Parallel to the creation of a language of deception, another hypothesis is that translation, especially automatic translation, could be a factory of fake news. Therefore, the analysis of texts generated by automatic

translators could be also decisive when comparing news spread in several languages, since a mistranslation of a news piece could be the cause of disinformation generation.

3.3 Automatic detection of fake news

After detecting the features that differentiate fake news from evidence-based news, machine learning could learn and automate the detection process in order to automatically identify fake news.

During this process, it is not only important to obtain a classification on whether a news piece is true, inaccurate or completely false, but also that the machine learns to justify the decision made.

4 Objectives of the research

In order to prove if the aforementioned hypothesis are viable, some main and secondary objectives have been set. The main objective of this thesis is the modelling of a language of deception by means of the linguistic, structural, emotional and content analysis for the automatic detection of fake news. This goal could be achieved by setting several specific objectives from different approaches:

- Study of the status of the issue: a research of the background and the progress made in NLP will be carried out in order to have a deep knowledge of the automatic detection of fake news.
- Corpus compilation: compilation of a trilingual corpus composed of fake news, true news and fact-checks in Spanish, English and French. News will belong to the health domain.
- Creation of a language model of deception: the model will be created on the basis of news structure and linguistic, psychosocial and emotional characteristics of the essential elements of news. The structural analysis will focus on the segmentation of news in parts and content elements; the linguistic analysis will comprise lexical, syntactic, semantic and orthotypographical approaches; the emotional analysis will focus on its influence in the viralization of fake news and the psychosocial study will tackle the issue of the influence of external factors in fake news, such as gender, social and cultural strata and discriminatory language.
- Cross-lingual analysis: study of the influence of the translation error in the creation of fake news. This objective will focus on analysing the mistakes detected in translations made by machine translators, which could lead to misinterpretations if they are not reviewed by translation experts (human translators).
- Development of an architecture that integrates the necessary HLT resources: the linguistic model created for fake news will be applied in machine learning by NLP experts.
- Validation of the proposal: the validation will be proved through experimentation and evaluation.

5 Methodology and experiments

The methodology adopted to address this project will firstly focus on an in-depth study of the background of fake news as well as the progress made in NLP in relation to this area to develop new ideas and thus work on the challenge of automatic detection of fake news.

In parallel, it is fundamental to compile a trilingual corpus composed of news in Spanish, English and French. The corpus will contain news belonging to the health domain, as it is a field that tends to suffer the consequences of fake news.

Another important step to take during this process is the contact with different organisations involved in the task of fake news detection in order to learn their know-how when detecting deception manually and to improve our system.

One of the main proposals of this work is to analyse each of the parts and elements that make up true and fake news in order to compare both and thus detect structural and linguistic features that could be key in the creation of fake news. We have already made significant progress in this regard by creating an annotation schema that helps to detect in which parts is more common to lie.

Two other approaches independent of the structural and content analysis are the translation and the psychosocial studies, which could also be decisive in detecting deception. On the one hand, the translation analysis would allow to know how translation, especially machine translation, influence the cre-

ation of fake news. Errors made by machine translators in news that will later be published without human review could be a means of disseminating false information. On the other hand, the psychosocial study would provide a different perspective on the context in which fake news is created and on the aim pursued. Surveys of both professionals on the field, psychologists and journalists, as well as people belonging to different sectors, will help to know the impact of fake news on the population, such as if there are specific social and cultural sectors that are more influenced or if there is a discriminatory language (sexist or racist) that can be key in detecting news only addressed to a particular sector.

Last but not least, a crucial stage is the creation of an architecture based on the defined language model that learns to identify the structure and the content elements and that allows not only to detect fake news, but also to obtain a justified interpretation of the decision taken during the automatic detection process.

As mentioned above, some progress has been made in the compilation and annotation tasks:

- Among this progress, it could be stressed the compilation of a corpus in Spanish composed of 200 news focused on the health domain. The corpus mixes fake and true news and they are classified in two topics: news related to general health problems, such as home remedies to "cure" diseases or to "improve" health, on the one hand, and news related to COVID-19, on the other (see Table 1).
- Besides the compilation of a corpus in a language that lacks resources in this domain, a new fine-grained annotation scheme has been proposed for machines to learn to distinguish and classify the different parts of news according to their veracity. The objectivity of a news piece depends on two key factors: neutrality and the inverted pyramid (Thomson, White, y Kitley, 2008). Moreover, news has been annotated from two different approaches: structure and content.
- In the case of the structure approach, Khan et al. (2018) states that "certain parts of news articles carry different level of useful information" and in the inverted pyramid hypothesis (see Figure 1) relevant information is placed at the top while the least important information is located at the bottom of a news piece (Zhang y Liu, 2016). The annotation was therefore focused on the analysis of the headline, subtitle, lead and conclusion. The objective was to know if annotating each part of a news piece separately we can detect where is the false information.
- The content approach is based on the journalistic technique known as 5W1H, a method consisting in answering six questions that are key to accurately communicate a story. These six questions are: WHAT, WHO, WHERE, WHEN, WHY and HOW. The structure can provide information of the part in which is more common to lie, but the content can allow to know which are the key elements in the creation of a lie.
- To the best of the authors' knowledge, there is no detection system that considers all the parts of a news piece separately, but systems that assign a single veracity value to the whole news piece. Our proposal is not only to detect disinformation, but also to exactly know where and how the lie is created. For that reason, a more accurate annotation scheme that considers all the structural and content elements of a news piece could be a suitable solution in automatic detection of fake news.
- The proposed architecture is composed of two layers: the structure layer (which divides the text according to the Inverted Pyramid hypothesis) and the veracity layer (which focus on determining the veracity of the 5W1H, as well as of the whole text). The first is divided into 2 phases (structural segmentation and essential content extractor) while the second consists of 3 phases (external enrichment, 5W1H veracity predictor and news article veracity predictor). Different experiments have been carried out in our corpus of 200 news. Firstly, an evaluation of the entire veracity layer with the 5W1H and its veracity value manually annotated (F1= 80%); secondly an evaluation of the phase 5 (veracity of the document) with the 5W1H manually

annotated and its veracity value (F1= 97%) and finally, an evaluation of the entire system which, using a plain text, is capable of carrying out all the phases automatically and obtaining a final result of F1= 67%, exceeding the baseline. In conclusion, our proposal increases the results, so it is a suitable solution.

Annotated news	Amount
Health - FAKE	47
Health - TRUE	53
Covid-19 - FAKE	48
Covid-19 - TRUE	52
Total	**200**

Table 1: Compilation of a corpus in Spanish composed of 200 annotated news focused on the health domain.

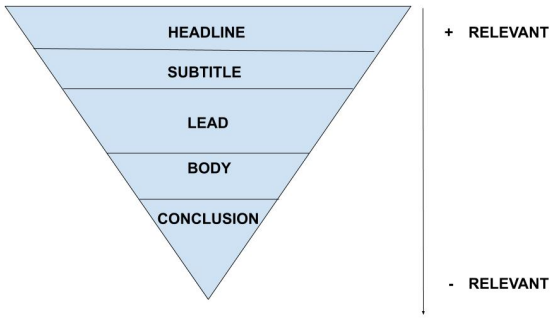

Figure 1: Inverted Pyramid Structure.

6 Specific elements of the research for discussion

This project aims to address several issues related to the automatic detection of fake news and the fight against disinformation:

6.1 Structural and content characteristics as a method to detect fake news

As explained throughout the article, fake news does not only contain false information, but also inaccurately or true information. The hyphotesis of this project is that a fine-grained annotation of each part of a news piece (headline, subtitle, lead, body and conclusion) and of each content element (5W1H) can help to detect false information, since the segmentation would allow to know in which part or element the lie is found.

6.2 Is the body a decisive part?

A compilation of a huge corpus is essential in this study, as a large number of examples are needed to detect the features of fake news language. As mentioned in Section 5, the inverted pyramid and the 5W1H method are essential for the annotation process. At the structural level, all the parts of a news piece have been annotated except one: the body. Is the body a crucial aspect in determining the veracity of a news piece?

Headline and lead are decisive in the annotation, since, according to the inverted pyramid hypothesis, they present the most relevant information. The annotation of the conclusion allows to know if the information is least important or not. But does the body comprise important information for the detection of fake news?

6.2.1 Experiment: the body annotation

As the body presents all the development of the story, it would be interesting to know if the 5W1H elements included in it are helpful when deciding the veracity of a news piece.

In order to answer the question about how the body affects when determining the veracity of a news item, an experiment will be carried out with an association that collaborates in the compilation and annotation of the corpus. In the experiment proposed, an annotator will annotate all the parts except the body part and two more annotators will do the same but also annotating the body part. Tasks will be rotated and the experiment will consist on detecting the veracity value with and without the body and on analysing if there is an agreement between the three annotators.

6.3 Linguistic analysis also needs extralinguistic information

Another future work would be to create a fact-checking module that will improve the results obtained by the linguistic analysis by adding external knowledge. This will also improve the machine learning process and the automatic detection of the veracity value.

Conclusion

Language has no rules and it is used in any situation and for any purpose. This is one of the reasons why detecting fake news by means of the language is so complicated. Language can be shaped in any way and a lie can be disguised as a true information or even to mix both true and false information in order to achieve a purpose.

However, our proposal is not impossible as news are not nothing other than a concatenation of words with a specific structure and meaning. Fake news seeks to influence society on a given issue and for a specific purpose and language could not only be its means of creation, but also its means of destruction, since a deep lexical, structural and content analysis can allow machines to learn to detect their linguistic weaknesses and to automatically differentiate truthful information from fake information.

Modelling language used in fake news and using Human Language Technologies could help to automatically detect fake news and justify the decision made. Translation, Sociology and Psychology also play an important role, since these three disciplines can allow to know the context and the causes of the fast dissemination of false information. All these tasks would not be possible without both manual and automatic detection, as linguistics and computer engineering must work together in order to make possible the human-machine relationship through the understanding and automatic generation of language. Artificial Intelligence, more specifically Natural Language Processing, needs Linguistics and vice versa to put an end to the era of disinformation.

Acknowledgements

This research work has been partially funded by Generalitat Valenciana through project "SIIA: Tecnologías del lenguaje humano para una sociedad inclusiva, igualitaria, y accesible" with grant reference PROMETEU/2018/089, by the Spanish Government through project RTI2018-094653-B-C22: "Modelang: Modeling the behavior of digital entities by Human Language Technologies", as well as being partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER) and the LIVING-LANG project (RTI2018-094653-B-C21) from the Spanish Government.

References

Alba-Juez, L. y J. L. Mackenzie. 2019. Emotion, lies, and "bullshit" in journalistic discourse: The case of fake news. *Ibérica*, (38).

Borges, L., B. Martins, y P. Calado. 2019. Combining similarity features and deep representation learning for stance detection in the context of checking fake news. *Journal of Data and Information Quality (JDIQ)*, 11(3):1–26.

Conroy, N. K., V. L. Rubin, y Y. Chen. 2015. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4.

Khan, S. U. R., M. A. Islam, M. Aleem, M. A. Iqbal, y U. Ahmed. 2018. Section-based focus time estimation of news articles. *IEEE Access*, 6:75452–75460.

Nycyk, M. 2015. The power gossip and rumour have in shaping online identity and reputation: A critical discourse analysis. *Qualitative Report*, 20(2).

Rashkin, H., E. Choi, J. Y. Jang, S. Volkova, y Y. Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. En *Proceedings of the 2017 conference on empirical methods in natural language processing*, páginas 2931–2937.

Saquete, E., D. Tomas, P. Moreda, P. Martinez-Barco, y M. Palomar. 2020. Fighting post-truth using natural language processing: A review and open challenges. *Expert Systems with Applications*, 141:112943.

Sukhodolov, A. P. y A. M. Bychkova. 2017. Fake news as a modern media phenomenon: definition, types, role of fake news and ways of counteracting it. , 6(2).

Thomson, E. A., P. R. White, y P. Kitley. 2008. "objectivity" and "hard news" reporting across cultures: Comparing the news report in english, french, japanese and indonesian journalism. *Journalism studies*, 9(2):212–228.

Uppal, A., V. Sachdeva, y S. Sharma. 2020. Fake news detection using discourse segment structure analysis. En *2020 10th In-*

ternational Conference on Cloud Computing, Data Science & Engineering (Confluence), páginas 751–756. IEEE.

Zhang, H. y H. Liu. 2016. Visualizing structural "inverted pyramids" in english news discourse across levels. *Text & Talk*, 36(1):89–110.