LCAD - UFES at FakeDeS 2021:
Fake News Detection Using Named Entity Recognition and Part-of-Speech Sequences

Marcos A. Spalenza[1], Leopoldo Lusquino-Filho[3], Felipe M. G. França[3],
Priscila M. V. Lima[2,3], and Elias de Oliveira[1]

[1] Postgraduate Program in Informatics (PPGI),
Federal University of Espírito Santo (UFES), Vitória-ES, Brazil.
[2] NCE, Tércio Pacitti Institute,
[3] Systems Engineering and Computer Science Program (PESC),
Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro-RJ, Brazil.
`marcos.spalenza@gmail.com elias@lcad.inf.ufes.br`

Abstract. News is fundamental to share interesting and relevant facts for public knowledge. However, unreliable sources produce fake and biased information, releasing content without proper fact-checking. The biased content attends to a massive disclosure on the internet and sociopolitical tendencies. Consequently, the identification of inaccurate news minimizes the damage to public entities. Therefore, against the misinformation, the fact-checking agencies investigate the trending news. Regarding the investigation, manual checking is slow and expensive. To filter these demands, we propose an automated method using linguistic components, supporting fake content identification. Our approach applies Machine Learning using POSTag+NER sequences. In the interdomain analysis, our method achieves 71% in the F1 measure for fake news detection.

Keywords: Fake News Detection · Natural Language Processing · Ensemble Learning · Named Entity Recognition

1 Introduction

The information acquired through the internet made the news agencies more dynamic. The duality of the news immediacy and the social networks' textual summaries increase the disclosure of unlawful, defamatory, threatening, false, or misleading content shared without verification [22]. However, these contents frequently origin from unreliable sources, generating content leaned to political or social entities.

The untruth inside news articles impacts the common knowledge, reflecting on public entities or social traditions [23]. The Covid-19 pandemic crisis emphasizes the misinformation problem. Fake news widespread worse the worldwide situation causing denial, self-medication and, political attacks [8, 13]. However, the countries study to treat the major unlawful consequences of fake information considering the freedom of expression rights. These concerns highlight the core problem of fake news, the subjective details that characterize a fact as truth or false [25].

Despite the subjectivity, we study methods to identify fake news among widespread topics using linguistic features. The linguistic features imply in document model references by textual sequential structures. In other words, our approach encompasses finding language models for false and true categories inside the news *corpora*.

This paper is presented in 5 sections. Section 2 describes some literature works on fake news detection. Section 3 present the POSTag+NER model, analyzing the documents using grammar sequences. Section 4 describes the fake news datasets and the obtained results. Finally, Section 5, present our conclusions and future works on fake news detection.

2   Related Works

The content production in the digital platforms is majority unverified, widely reproduced and, continuously modified. The production of fake content is unbalanced given the demand for verification and filtering [21]. These high demands are a consequence of the disclosure of fake information through social networks and unreliable news sites [3].

The propagation of fake news encompasses diverse factors. The media, the source, the propagation, the references and, images or textual content compose the news articles data. [6, 14]. Some works highlight the importance of metadata to classify an article as fake or true [2, 27]. The metadata traces the false publication likelihood through the replication chain, the user comments, the cross-references and, the news agency general reliability. Although, the contents are the main evidence of factual and non-factual knowledge. Regarding the news articles' contents, the linguistic features are the most studied approach to investigate the untruthful content dissemination [4].

In linguistics, to support the verification process and restrain misinformation, the fake news detection systems seek factual and non-factual contents [5, 20]. The methods comprise searching for similar structures on the text, sentence descriptors, writing models and, linguistic sequences [10, 17]. Using the linguistic descriptors, the models extract the news' writing patterns. These descriptors comprise syntax, polarity, grammar and, readability levels.

The content-based methods analyze the textual and visual information to detect fake content [26]. The textual analysis includes identifying equivalent propositions, weighting potential incoherent words or sequences. Furthermore, the modified, symmetric or, related textual features are a vulnerability on basic

content-based detection [15, 28]. In other words, the systems fact-checking have to be robust to avoid textual bias. In this perspective, the recent studies aim to advance in the construction of linguistic models, to improve the assessment of writing coherence, information quality and, inter-domain adaptability [10, 20].

3   Model

Our approach includes identifying the language models for factual and non-factual articles, recognizing the specific linguistic structures. The language modeling aims to identify *fake* and *true* classes through writing patterns [24]. However, these patterns represent proper writing styles, defined by textual sequences.

In this paper, our analysis using Part-Of-Speech Tags (POSTags) to compare the news articles through the grammar sequences. We aim to produce a system that recognizes and learns to evaluate the coherence of the writing. Therefore, we expect to find in fake categorized articles some incoherent, biased, repetitive and, incorrect language format. We present an example to illustrate the process to generate the linguistic models. In the first step, the POSTagger, apply the language models to convert the words in their grammatical references.

Example

```
Luego de que se revelara que el ex alcalde de Cuernavaca y
ahora aspirante de la gubernatura de Morelos, Cuauhtémoc Blanco,
aún sigue registrado como jugador del Club América y en breve
participará en un partido, el director técnico de la selección
mexicana, Juan Carlos Osorio, reveló que Cuauhtémoc será uno
de los futbolistas que serán convocados para representar a
México en el Mundial de Rusia del *NUMBER*.
```

Part-of-Speech

```
ADV ADP SCONJ PRON VERB SCONJ DET ADJ NOUN ADP PROPN CCONJ
ADV NOUN ADP DET NOUN ADP PROPN PUNCT PROPN PROPN PUNCT ADV
VERB ADJ SCONJ NOUN ADP PROPN PROPN CCONJ ADP NOUN VERB ADP
DET NOUN PUNCT DET NOUN ADJ ADP DET NOUN ADJ PUNCT PROPN PROPN
PROPN PUNCT VERB SCONJ PROPN AUX PRON ADP DET NOUN PRON AUX
VERB ADP VERB ADP PROPN ADP DET PROPN ADP PROPN ADP SYM PROPN
SYM PUNCT
```

In the example, we observe the news original text and the words' grammar tags. Afterward, we apply Named Entity Recognition (NER) to identify the entities that compose the news article and classify their semantic role. Using the NER, we search the fake news targets within the textual components looking for name references, such as politicians, organizations, locations or, public people. The entities' classification include four categories: *Person*, *Organization*, *Local* and, *Miscellaneous*. In the original text sequence, together the POSTag and NER transform the sentences in the grammar functions and the specific name semantics. The second step detail the detection and classification of named entities within the selected example.
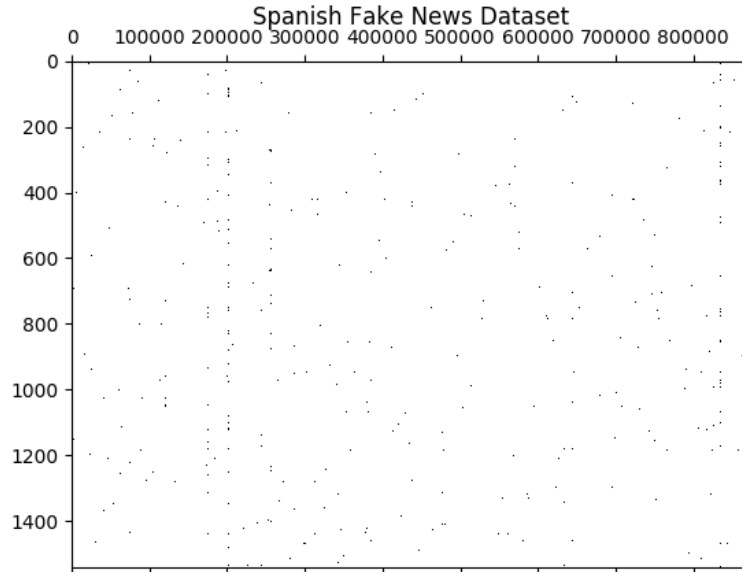
Named Entity Recognition

```
O O O O O O O O O LOC O O O O O O O PER O PER PER O O O O
O O O ORG ORG O O O O O O O O O O O O O O O PER PER PER O
O O PER O O O O O O O O O O O PER O O MISC MISC MISC O O NUM
O O
```

```
Cuernavaca (LOC)
Morelos (PER)
Cuauhtémoc (PER) Blanco (PER)
Club (ORG) América (ORG)
Juan (PER) Carlos (PER) Osorio (PER)
Cuauhtémoc (PER)
México (PER)
Mundial (MISC) de (MISC) Rusia (MISC) NUMBER (NUM)
```

Applying the POSTags+NER sequences [24], the system analyzes the words only by their functions on the sentence. Considering the words individually, the article's factual content not necessarily is lined to be categorized as *fake* or *true*. Therefore, in this perspective, the words' frequencies probably are a bias to categorize an article by its content [15]. Our example outlines the POSTag and NER combination at the final of the preprocessing.

POSTag+NER

```
ADV ADP SCONJ PRON VERB SCONJ DET ADJ NOUN ADP LOC CCONJ ADV
NOUN ADP DET NOUN ADP PER PUNCT PER PER PUNCT ADV VERB ADJ
SCONJ NOUN ADP ORG ORG CCONJ ADP NOUN VERB ADP DET NOUN PUNCT
DET NOUN ADJ ADP DET NOUN ADJ PUNCT PER PER PER PUNCT VERB
SCONJ PER AUX PRON ADP DET NOUN PRON AUX VERB ADP VERB ADP
PER ADP DET MISC MISC MISC ADP NUM PUNCT
```

To sum up, the POSTag+NER is a text preprocessing method to organize the data in vectors of linguistic patterns. These patterns include to learn the news targets, language, correctness and, factual structures. However, the model consists in a grammar sequences using documents' *n-grams*. Additionally, we recognize the entities to replace the original grammar tag. The entities aims to identify the news characters using a semantic description and integrating the sequence model.

At the vectorization step, we apply the Term Frequency (TF) to generate the high-dimensional and sparse document vectors, using 3 up to 7-grams sequences. Figure 1 presents the sparse document matrix.

Figure 1 shows the matrix containing 1,543 document vectors and 874,584 features presenting only 0.2369% filled area. Despite the class subjectivity concerning the document's content, we highlight the features' low occurrence (sparsity) and the coverage of linguistic structures (high-dimensionality). We test four classifiers to analyze the linguistic modeling from different perspectives: Support Vector Machines (SVM), Random Forest (RF), Gradient Boosting (GB) and, Wilkie, Stonham & Aleksander's Recognition Device (WiSARD).

Fig. 1. The POSTag+NER features' dispersion in the sparse matrix.

The SVM evaluates the document dispersion in kernels. The kernels define feature threshold hyperplanes to classify the unviewed samples [18]. The RF identifies features' threshold and classifies the samples through multiple decision trees [7]. The GB is a combination of decision tree models through differentiable loss functions, reducing the global error for each weak learner on iterative improvements [11]. Finally, the WiSARD [1], Weightless Neural Network using a binary thermometer [9] data encoding and classifying the samples by matching the class binary patterns.

The linguistic pre-trained models for POSTagger and NER methods are provided from *spaCy*[4] and the classifiers from *scikit-learn*[5] and *wisardpkg*[6]. Through different classifiers, we evaluate the learning of the linguistic patterns to produce the class models.

4 Experiments and Results

The conducted experiments evaluated the POSTag+NER approach for Spanish fake news detection. To perform false content detection, the systems need to

---

[4] https://spacy.io/

[5] https://scikit-learn.org/

[6] https://github.com/IAZero/wisardpkg/

identify the untruth among different data sources. However, it is fundamental to the systems adapt to multiple domains, sources, and contents.

In this experiment, we tested the system inter-domain classification. The training dataset contains 917 news in Spanish [19]. The data contains 491 real and 480 fake articles about science, sport, economy, education, entertainment, politics, health, security, and social domains. The true class was manually tagged from reliable news agencies and the fake class from specialized verification sites.

The test data contains 572 news articles from the Covid-19 pandemic, 286 true and 286 fake news [12]. The dataset includes news articles from multiple Ibero-American countries. The challenge is to recognize relations between the documents considering the dissimilar themes [16]. Additionally, the news articles present the regional language adaptations from each country. Furthermore, our approach aims to identify linguistic models and the factual structure for each category, despite of its textual content.

To evaluate the system, we apply the F1 score to measure the detection performance. The F1 is the ratio between the precision and recall scores. On the fake news detection, the evaluation focus is the binary categorization for positive class identification. Besides that, we executed the system ten times to collect the standard deviation of classifiers' performance. Table 1 discriminates the classifiers' maximum and average results including the standard deviation.

Table 1. The results achieved on the Spanish Fake News Detection Task

IberLEF - Fake News Detection Task				
Train: Spanish Fake News Dataset			*Test:* FakeDeS Covid-19 Dataset	
		Classifiers		
		GB	RF	SVM	WiSARD
F1 Score Max. (%)		71.14	67.41	64.81	64,01
Avg. (%)		70.83	65.63	64.81	63,28
Std.Dev.		0.1807	0.9660	0.0000	0,4652

Table 1 highlights the performance for the tested classifiers at the *IberLEF Spanish Fake News Detection Task*. Regarding the different approaches, in an inter-domain perspective, the four methods present good classification capabilities. Through the WiSARD, the adaptability for inter-domain classification was averaging 63% in the F1 score, using 3-bits thermometer encoding and 50 on address size. This case, specifically, indicates an insufficient generalization on high-dimensional binary models. The training vectors, composed mostly of zero values, present low compatibility to the test samples' grammar sequences. As a consequence, the test vectors are similar to both binary class patterns.

The SVM mounts the kernels identifying the features' threshold. The train samples outline a regular SVC kernel zone, reinforcing the relevance of key features. Although, these kernels do not outline an efficient class area. Despite that, the observed performance was slightly better than WiSARD, approximately 65% in F1 score. In general, the SVM and WiSARD have good performance. Accord-

ing to the results, the ensemble methods design a more robust and efficient classification.

The ensemble training step of RF and GB produces complex class models, merging robust classifiers. On one hand, the RF combines 250 weak learners using feature analysis to identify relevant information bias in spite of the high-dimensional and sparse samples. On the other hand, the GB reduces the global error by combining 1000 weak learners through a learning rate of 0.03. The RF classification reaches 67% and GB classification reaches 71% in the F1 score.

Analyzing the results, we observe a higher performance from the GB in relation to the other classifiers. Despite the classifier, the learning and the performance of the POSTag+NER linguistic model present good results in the inter-domain experiments. The results confirm the GB information acquisition presenting over 70% F1 score and low standard deviation.

5 Conclusion

Fake news is a crescent problem on social, economical and, political aspects. The automatic detection of false content reduces public replication and misinformation. In this problem, we propose the application of Named Entity Recognition in addition to Part-Of-Speech tag sequences.

The experiments establish an inter-domain challenge, evaluating the model adequacy between a regular fake news dataset and a thematic Covid-19 fake news dataset. The POSTag+NER results are similar to other approaches, presenting high-level performances. To improve this work, the next steps comprise analyze the most relevant features to iteratively create better models and identify the features' textual entailment related to the non-factual information. In addition, we expect to analyze and enhance inter-domain detection.

Acknoledgements

References

1. Aleksander, I., Thomas, W., Bowden, P.: WISARD· A Radical Step Forward in Image Recognition. Sensor Review (1984)
2. Alhindi, T., Petridis, S., Muresan, S.: Where is Your Evidence: Improving Fact-checking by Justification Modeling. In: Proceedings of the First Workshop on Fact Extraction and VERification (FEVER). pp. 85–90. Association for Computational Linguistics (Nov 2018)
3. Apuke, O.D., Omar, B.: Fake News and COVID-19: Modelling the Predictors of Fake News Sharing Among Social Media Users. Telematics and Informatics **56**, 101475 (2021)

4. Beer, D., Matthee, M.: Approaches to Identify Fake News: A Systematic Literature Review. Integrated Science in Digital Age pp. 13–22 (2020)
5. Bonet-Jover, A., Piad-Morffis, A., Saquete, E., Martínez-Barco, P., Ángel García-Cumbreras, M.: Exploiting Discourse Structure of Traditional Digital Media to Enhance Automatic Fake News Detection. Expert Systems with Applications **169**, 114340 (2021)
6. Bozarth, L., Budak, C.: Toward a Better Performance Evaluation Framework for Fake News Classification. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 14, pp. 60–71 (May 2020)
7. Breiman, L.: Random Forests. Machine Learning **45**(1), 5–32 (2001)
8. Bridgman, A., Merkley, E., Loewen, P.J., Owen, T., Ruths, D., Teichmann, L., Zhilin, O.: The Causes and Consequences of COVID-19 Misperceptions: Understanding the Role of News and Social Media. Harvard Kennedy School Misinformation Review **1**(3) (2020)
9. Buckman, J., Roy, A., Raffel, C., Goodfellow, I.: Thermometer Encoding: One Hot Way To Resist Adversarial Examples. In: International Conference on Learning Representations (2018)
10. Choudhary, A., Arora, A.: Linguistic Feature Based Learning Model for Fake News Detection and Classification. Expert Systems with Applications **169**, 114171 (2021)
11. Friedman, J.H.: Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics **29**(5), 1189–1232 (Oct 2001)
12. Gómez-Adorno, H., Posadas-Durán, J.P., Bel-Enguix, G., Porto, C.: Overview of FakeDeS Task at Iberlef 2020: Fake News Detection in Spanish. Procesamiento del Lenguaje Natural **67**(0) (2021)
13. Hartley, K., Vu, M.K.: Fighting Fake News in the COVID-19 Era: Policy Insights from an Equilibrium Model. Policy Sciences **53**(4), 735–758 (2020)
14. Kaliyar, R., Goswami, A., Narang, P.: A Hybrid Model for Effective Fake News Detection with a Novel COVID-19 Dataset. In: Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART). vol. 2, pp. 1066–1072. INSTICC, SciTePress (2021)
15. Monteiro, R.A., Santos, R.L.S., Pardo, T.A.S., de Almeida, T.A., Ruiz, E.E.S., Vale, O.A.: Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results. In: Computational Processing of the Portuguese Language. pp. 324–334. Springer International Publishing, Cham (2018)
16. Montes, M., Rosso, P., Gonzalo, J., Aragón, E., Agerri, R., Álvarez-Carmona, M.A., Mellado, E.A., Carrillo-de Albornoz, J., Chiruzzo, L., Freitas, L., Adorno, H.G., Gutiérrez, Y., Zafra, S.M.J., Lima, S., Plaza-de Arco, F.M., Taulé, M. (eds.): Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021). CEUR Workshop Proceedings, CEUR-WS, Spain (2021)
17. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R.: Automatic Detection of Fake News. In: Proceedings of the 27th International Conference on Computational Linguistics. vol. 17, pp. 3391–3401. Association for Computational Linguistics, Santa Fe, New Mexico, USA (Aug 2018)
18. Platt, J., et al.: Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. Advances in Large Margin Classifiers **10**(3), 61–74 (1999)
19. Posadas-Durán, J.P., Gómez-Adorno, H., Sidorov, G., Escobar, J.J.M.: Detection of Fake News in a New Corpus for the Spanish Language. Journal of Intelligent & Fuzzy Systems **36**(5), 4869–4876 (Jan 2019)

20. Rashkin, H., Choi, E., Jang, J.Y., Volkova, S., Choi, Y.: 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. pp. 2931–2937. Association for Computational Linguistics, Copenhagen, Denmark (Sep 2017)

21. Shao, C., Ciampaglia, G.L., Varol, O., Yang, K.C., Flammini, A., Menczer, F.: The Spread of Low-Credibility Content by Social Bots. Nature Communications **9**(1), 1–9 (2018)

22. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake News Detection on Social Media: A Data Mining Perspective. SIGKDD Explorations Newsletter **19**(1), 22–36 (Sep 2017)

23. Shu, Kai and Wang, Suhang and Lee, Dongwon and Liu, Huan: Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements, pp. 1–19. Springer International Publishing, Cham (2020)

24. Spalenza, M.A., Oliveira, E., Lusquino-Filho, L.A.D., Lima, P.M.V., França, F.M.G.: Using NER+ML to Automatically Detect Fake News. In: 20th. International Conference on Intelligent Systems Design and Applications (ISDA). vol. 20, pp. 1176–1187 (Dec 2020)

25. Vieira, L.L., Jeronimo, C.L.M., Campelo, C.E.C., Marinho, L.B.: Analysis of the Subjectivity Level in Fake News Fragments. In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia '20). pp. 233—-240. Association for Computing Machinery, São Luís, Brazil (2020)

26. Zhou, X., Wu, J., Zafarani, R.: $SAFE$: Similarity-Aware Multi-modal Fake News Detection. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. pp. 354–367. Springer (2020)

27. Zhou, X., Zafarani, R.: Network-based Fake News Detection: A Pattern-Driven Approach. ACM SIGKDD Explorations Newsletter **21**(2), 48–60 (2019)

28. Zhou, Z., Guan, H., Bhat, M., Hsu, J.: Fake News Detection via NLP is Vulnerable to Adversarial Attacks. In: Proceedings of the 11th International Conference on Agents and Artificial Intelligence - ICAART. vol. 2, pp. 794–800. INSTICC, SciTePress, Prague, Czech Republic (2019)