

Multilingual Fake News Detection with Satire

Gaël Guibon^{a,b}, Liana Ermakova^c, Hosni Seffih^{d,g}, Anton Firsov^e and Guillaume Le Noé-Bienvenu^f

^aLTCI, Télécom Paris, Institut Polytechnique de Paris

^bInnovation Recherche, SNCF, Saint-Denis, France

^cHCTI - EA 4249, Université de Bretagne Occidentale, Brest, France

^dGEOLSemantics, 12 avenue Raspail, 94250 Gentilly

^eLIASD - IUT de Montreuil - Université Paris 8, 140 Rue de la Nouvelle France, 93100 Montreuil

^fKnoema Corporation Perm, Russia

^gCNRS, IRISA, Rennes, France

Abstract

This is an extended abstract from a paper published at CICLing 2019 [1].

Keywords

fake news, deception, satire, artificial intelligence, machine learning, natural language processing

1. Extended Abstract

Humor is the way to provoke laughter and comes in different flavors. In this paper we only consider the understanding of one of them: the satire type of humor. According to Cambridge dictionary, satire is “a way of criticizing people or ideas in a humorous way, especially in order to make a political point, or a piece of writing that uses this style”. Moreover, satire often consists of reproducing a slightly exaggerated version of the criticized object, which makes the distinction between news (fake or trusted) and satire more difficult.

Due to the rise of fake news detection research topics, satire is becoming a type of humor often considered by Natural Language Processing or Information Retrieval researchers. At first, satire detection from trusted news only was done using classic classification methods mixing SVMs, feature scaling and different lexical and semantic feature types, focusing on semantic validity [2]. Satire has also been considered as harmful as hoaxes if the context is not provided, hence it has been categorized as a part of humorous fakes [3]. The satire detection is often done using tweets labeled by Twitter accounts’ reliability [2, 4, 5]. This approach was used for satire detection in Spanish written tweets using multiple features such as sentiment and WordNet Spanish [6] to capture synonyms and ambiguity [4]. Satire detection was also done in tweets along with suspicious news, hoaxes, click bait and propaganda, yielding to the satire tweets being the most different from hoaxes and propaganda and containing more bias markers than propaganda and fake news. Moreover, while classifying them, syntax and grammar features have little effect on the model’s performance [5]. More recent works used CNN [7] and attentive

Words/Machines-2021, March 05, 2021, Brest, France



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Bi-Directional GRU [8] or attention mechanism [9] over linguistic features to detect satirical Fake News [10, 11], using a True versus Satire news classification data set [11]. The comparison of several supervised algorithms for Fake News detection including satire has been done using BuzzFeed political news data set and yielded the best results using a decision tree algorithm [12]. In this paper published at CICLing 2019 [1], we focused on detecting Satire, Trusted and Fake News altogether while varying multiple features types.

Our approach is multilingual considering French and English as one using news articles, transcribed YouTube videos, and related metadata, all dedicated on vaccination topics and provided by the company Storyzy¹ specialized in Fake News detection. In this data set, the satire news and the multi-modal sources creates a noisy context for Fake News detection by considering this classification task a multi-class one. We consider this classification task by mixing corpus exploration strategies to obtain features, with supervised learning strategies.

Corpus exploration by text resemblance and domain type detection. Text resemblance score is a similarity ratio considered to give additional hints about the differences between Satire, Trusted and Fake News. Domain type detection is a matching comparison using a list of famous websites tagged as Trusted, Fake or Satire. Both text resemblance and domain type detection were computed on a web snapshot using the chatNoir API [13].

Supervised learning. We conducted multi-class classification experiments by mixing previous corpus exploration features with common data representations from simple TF-IDF vectors to word embeddings and sub-word embeddings. These representations are then fed to several types of classifiers: kernel based (SVM), decision tree based (LGBM) and neural network (CNN). We obtained the highest macro F1-scores and the best satire detection while using the LGBM classifier which precludes the same conclusion obtained one year later by Osbay *et al.* [12].

The main contribution of this work is the balance we obtained in the multi-class classification task while obtaining high scores on the satire class despite the data set being very unbalanced and the satire news under represented. Because we compare multiple approaches to detect not only Satire News, but also, at the time and with the same model, Fake News and Trusted News, this work is still relevant in regards to today's literature. Moreover, trying to maximize the overall classification quality over focusing on one specific task goes in the opposite direction to most related work considering satire detection. Our approach by feature stacking shows the opposite conclusion compared to *Burfoot et al.* [2] as scaling did not help but naive stacking did. We consider stacking as voluntary inducing redundancy in data representation to better help the classifier. This still shows limits regarding the high dimensional representation for each entry. However, LGBM being a decision tree based algorithm, feature importance score can help reduce the dimension to scale the experiment.

References

- [1] G. Guibon, L. Ermakova, H. Seffih, A. Firsov, G. Le Noé-Bienvenu, Multilingual fake news detection with satire, in: CICLing: International Conference on Computational Linguistics and Intelligent Text Processing, 2019.

¹<https://storyzy.com/>

- [2] C. Burfoot, T. Baldwin, Automatic satire detection: Are you having a laugh?, in: Proceedings of the ACL-IJCNLP 2009 conference short papers, 2009, pp. 161–164.
- [3] V. L. Rubin, Y. Chen, N. K. Conroy, Deception detection for news: three types of fakes, Proceedings of the Association for Information Science and Technology 52 (2015) 1–4.
- [4] F. Barbieri, F. Ronzano, H. Saggion, Is this tweet satirical? a computational approach for satire detection in spanish, Procesamiento del Lenguaje Natural (2015) 135–142.
- [5] S. Volkova, K. Shaffer, J. Jang, N. Hodas, Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), volume 2, 2017, pp. 647–653.
- [6] A. Fernández-Montraveta, G. Vázquez, C. Fellbaum, The spanish version of wordnet 3.0, Text resources and lexical knowledge 8 (2008) 175–182.
- [7] K. Fukushima, Neural network model for a mechanism of pattern recognition unaffected by shift in position-neocognitron, IEICE Technical Report, A 62 (1979) 658–665.
- [8] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder-decoder for statistical machine translation, 2014. [arXiv:1406.1078](https://arxiv.org/abs/1406.1078).
- [9] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, arXiv preprint [arXiv:1409.0473](https://arxiv.org/abs/1409.0473) (2014).
- [10] S. De Sarkar, F. Yang, A. Mukherjee, Attending sentences to detect satirical fake news, in: Proceedings of the 27th international conference on computational linguistics, 2018, pp. 3371–3380.
- [11] F. Yang, A. Mukherjee, E. Dragut, Satirical news detection and analysis using attention mechanism and linguistic features, arXiv preprint [arXiv:1709.01189](https://arxiv.org/abs/1709.01189) (2017).
- [12] F. A. Ozbay, B. Alatas, Fake news detection within online social media using supervised artificial intelligence algorithms, Physica A: Statistical Mechanics and its Applications 540 (2020) 123174.
- [13] J. Bevendorff, B. Stein, M. Hagen, M. Potthast, Elastic chatnoir: Search engine for the cluweb and the common crawl, in: G. Pasi, B. Piwowarski, L. Azzopardi, A. Hanbury (Eds.), Advances in Information Retrieval, Springer International Publishing, Cham, 2018, pp. 820–824.