Fake News Detection for Portuguese with Deep Learning

Lígia Iunes Venturott[1,2][0000−0002−6339−4700] and Ruslan Mitkov[2][0000−0003−2522−066X]

[1] New Bulgarian University, Sofia, Bulgaria
[2] University of Wolverhampton, Wolverhampton, UK

Abstract. The exponential growth of the internet and social media in the past decade gave way to the increase in dissemination of false or misleading information. Since the 2016 US presidential election, the term "fake news" became increasingly popular and this phenomenon has received more attention. In the past years several fact-checking agencies were created, but due to the great number of daily posts on social media, manual checking is insufficient. Currently, there is a pressing need for automatic fake news detection tools, either to assist manual fact-checkers or to operate as standalone tools. There are several projects underway on this topic, but most of them focus on English. This research-in-progress paper discusses the employment of deep learning methods, and the development of a tool, for detecting false news in Portuguese. As a first step we shall compare well-established architectures that were tested in other languages and analyse their performance on our Portuguese data. Based on the preliminary results of these classifiers, we shall choose a deep learning model or combine several deep learning models which hold promise to enhance the performance of our fake news detection system.

Keywords: Fake news detection · Deep learning · Portuguese.

1 Introduction

The term "fake news" is relatively new, having emerged in the 19th century. Before the 18th century the word "fake" was seldom used as an adjective and the expression "false news" was more common. According to Google Trends, the search for the term has significantly increased since October/November 2016, coinciding with the 2016 US presidential elections.

Even though the term gained popularity only in the past few years, the phenomenon of fake news itself is not new. Misinformation, disinformation, propaganda, conspiracy theories and others have always existed. However, the spread of this kind of content has exponentially grown with the new communication technology.

As a result of the rapid growth of the internet in the past decades, as well as the growth of social media, false or misleading information have been spreading at an alarming rate. Social media introduced a new kind of public space, one

Translation and Interpreting Technology Online, pages 149–153,
July 05-07, 2021.
https://doi.org/10.26615/978-954-452-071-7_016

where every individual has the opportunity to voice their opinion and potentially be heard by any other individual with internet access. The growth of social media and the lack of control of online content are major contributing factors to the phenomenon of fake news. Nowadays any person or organisation can create a social media profile and disguise as a professional news outlet. These actors provide misleading information which pretends to be reliable news.

Fake news also tend to be more appealing than true stories. Tweets containing false information reach 1500 persons six times faster than tweets containing real information [13]. According to Garner's prediction, "By 2022 most people in mature economies will consume more false information than true information"[1].

Several fact-checking agencies have surged as a response to the growth in fake news dissemination. However, manual checking of news is clearly insufficient when we consider the volume of posts in any of the major social media platforms [9]. Only Twitter, for example, had 340 million users and 500 million tweets per day on December 2020 [8].

While most work on fake news detection has been done for English, the topic has been scarcely researched for Portuguese and this ongoing study seeks to fill in this gap. Our objective is not only to develop additional fake detection tools for Portuguese, but to employ latest Deep Learning techniques in order to establish whether they will enhance the state-of-the-art of fake detection for Portuguese.

The rest of the paper is structured as follows. Section 2 surveys the work on fake detection for Portuguese so far. Section 3 of this work-in-progress poster paper outlines the envisaged methodology and briefly touches on the planned evaluation.

2 Related Work

Reis et al.[9] analysed several classification algorithms that use supervised learning. They use a dataset consisting of 2282 BuzzFeed articles about the 2016 US presidential election labelled by journalists. The authors explain three types of information can be extracted from news: content features, source features and environmental features, and extract several features of each of these categories.

Several classification algorithms were selected: k-Nearest Neighbours (KNN), Naive Bayes (NB), Random Forests (RF), Support Vector Machine with RBF kernel (SVM), and XGBoost (XGB). These models were used to evaluate the discrimination power of the extracted features. While it would have been promising to experiment with deep learning techniques, the authors opted for hand crafted features. The manual extraction of these features is probably time consuming, even with the help of automatic tools.

Silva et al.[10] evaluated different detection methods on the Fake.BR dataset[7], a dataset of fake news in Portuguese. The authors go about their study in two different ways: by extracting linguistic features from the data and by employing vector representations. They chose three types of vector representations: bag-of-words with TF-IDF, Word2Vec and Fast Text, and used pre-trained Word2Vec and Fast Text vectors. The idea was to compare what would perform better:

hand-crafted features or automatically extracted features. Several classification algorithms were selected in order to evaluate the representations: logistic regression (LR), support vector machines (SVM), decision trees (DT), random forest (RF), bootstrap aggregating (bagging) and adaptive boosting (Ada-Boost). The experiments with linguistic based features showed that these features are good enough to detect more than 90% of the fake news. This result is very interesting, because hand-crafted features do not require complex classifiers, meaning that a detection algorithm could run on even the simplest of devices. In the experiments with vector representations, the authors conclude that the bag-of-words model delivered better results than Fast Text or Word2Vec. The best F-measure obtained with Word2Vec and FastText were 0.893 and 0.897, respectively, while the best F-measure with BoW was 0.971. The authors hypothesise that, since the pre-trained vectors were trained on well-written texts such as Wikipedia and Google News, these vectors were not the best fit to represent fake news, which contains noise, such as incorrect spelling and slang.

3 Methodology

3.1 Datasets

As supervised learning will be employed in this project, a dataset of fake and real news will be needed to train the algorithm. So far we have identified three datasets available online.

The first, Fake.BR [7], is a balanced dataset containing 7,200 news, of which 3,600 classified as fake and 3,600 classified as true. The second, FACTCK.BR, is unbalanced and contains 1,309 news in total, of which 943 true, 246 half-true and 120 false. The above dataset does not offer explicit texts, only the url from where the news was taken, which means that in order to use it we would need to add another step of web scraping. The third, Boatos.org is available on Kaggle[3] and contains 1,900 WhatsApp messages proved to be fake news by fact-checking agencies. Unfortunately, it does not contain messages without false information, so we cannot use this dataset alone to train the algorithms.

If the aforementioned datasets turn out not to be suitable for our study, we shall compile a small annotated corpus of fake news in Portuguese.

3.2 Preprocessing

A preprocessing pipeline will be also implemented in order to minimise the noise in the data. Normally the pipeline for text preprocessing consists of: (1) Tokenisation, (2) Normalisation (lower case, remove accents and special characters, convert to ascii), (3) Lemmatisation or Stemming, (4) Stop word removal and (5) Numeralisation. Other steps may be added if necessary.

[3] https://www.kaggle.com/rogeriochaves/boatos-de-whatsapp-boatosorg

3.3 Classifiers

In this project we shall use deep learning techniques to detect fake news. For this purpose, we will evaluate and compare different deep architectures and identify additional ones, if needed, that hold promise for high performance on this task.

LSTM There are several deep learning architectures available, but because we are dealing with text, which is a type of sequential data, we will start our experiments with the LSTM (Long Short-Term Memory)[6]. LSTMs were introduced as an improvement to regular Recurrent Neural Networks (RNN). An RNN analyses each word in a sentence at a time. At each time step the layer analyses one word and generates an output. This output is then used as an additional input for the next time step with the next word. By using this mechanism, when analysing a word in a sentence the layer has access to information about the previous word and, recursively, about every word that occurred before. Simple RNNs are not suited for real-world applications due to the vanishing/exploding gradient problem [6]. In the LSTM, the simple summation cell is substituted for a memory block with 3 gates. These gates allow the information to be stored for more time steps, what remedies the problem of vanishing gradients[5].

Attention Based on our literature review, we believe that the performance might be boosted by using attention mechanisms[2]. The attention mechanism was originally crated in the context of machine translation. It allows the decoder to focus on important areas of the source sentence during the generation of the target sentence. The attention mechanism can also be used in classification tasks, and it might help improve the performance of the LSTM.

BERT The Transformer architecture[12] was also created for machine translation. It uses layers of multi-head self-attention mechanism and simple feed-forward networks to build the encoder and decoder. BERT is a language representation model based on transformers[4]. It can be pre-trained on unlabelled data, and then fine-tuned for specific tasks. There is already a pre-trained BERT model for Brazilian Portuguese[11]. We plan to test this model with fine-tuning and compare it with the other approaches.

Based on the preliminary results of these classifiers, we shall choose a deep learning model or combine several deep learning models which hold promise to enhance the performance of our fake news detection system.

3.4 Evaluation

We will evaluate our fake news detection system using standard evaluation metrics such as Accuracy and F1-score. We will compare our system with other existing fake detection systems for Portuguese, such as the ones mentioned in Section 2. In order to report statistical significance, we plan to use the Wilcoxon signed-rank test[3]. In addition to this intrinsic evaluation, we also envisage extrinsic evaluation where users evaluate the efficiency of the tool to be developed.

References

1. Gartner reveals top predictions for it organizations and users in 2018 and beyond (Oct 2017), https://www.gartner.com/en/newsroom/press-releases/2017-10-03-gartner-reveals-top-predictions-for
-it-organizations-and-users-in-2018-and-beyond
2. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
3. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. **7**, 1–30 (Dec 2006)
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding (2019)
5. Graves, A.: Long short-term memory. In: Supervised sequence labelling with recurrent neural networks, pp. 37–45. Springer (2012)
6. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. Neural Computation **9**(8), 1735–1780 (11 1997). https://doi.org/10.1162/neco.1997.9.8.1735, https://doi.org/10.1162/neco.1997.9.8.1735
7. 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 (2018)
8. Omnicore: Omnicore. https://www.omnicoreagency.com/twitter-statistics/ (2020), accessed 05/05/2021
9. Reis, J.C.S., Correia, A., Murai, F., Veloso, A., Benevenuto, F.: Supervised learning for fake news detection. IEEE Intelligent Systems **34**(2), 76–81 (2019). https://doi.org/10.1109/MIS.2019.2899143
10. Silva, R.M., Santos, R.L., Almeida, T.A., Pardo, T.A.: Towards automatically filtering fake news in portuguese. Expert Systems with Applications **146**, 113199 (2020). https://doi.org/https://doi.org/10.1016/j.eswa.2020.113199, https://www.sciencedirect.com/science/article/pii/S0957417420300257
11. Souza, F., Nogueira, R., Lotufo, R.: BERTimbau: pretrained BERT models for Brazilian Portuguese. In: 9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23 (to appear) (2020)
12. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. arXiv preprint arXiv:1706.03762 (2017)
13. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**(6380), 1146–1151 (2018). https://doi.org/10.1126/science.aap9559, https://science.sciencemag.org/content/359/6380/1146