Fake News and Sarcasm, what is the limit of a critic and what is intentionally fake?

Fernando Cardoso Durier da Silva[1], **Ana Cristina Bicharra Garcia**[1]

[1] Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

{fernando.durier,cristina.bicharra}@uniriotec.br

Abstract. *Nowadays it is hard to distinguish what is a fake information made to mislead, cause havoc and hysteria, than a true critic that has only the intention to highlight a social problem or an abnormality of some sort.*
In order to diminish the number of false positives in fake news detection, we experimented two neural network models trained by a combined set of true, fake and sarcastic news in order to test how accurate would be our model.
This paper has the goal to propose future steps of an ongoing experiment and discuss the usage of collaboration aided by machines to gather such data to the models.

1. Introduction

Nowadays, we experiment life in physical and virtual worlds. In the physical world, we interact physically with the world and other people. Meanwhile in the virtual world we also create experiences through social medias by consuming information and producing content to express ourselves in that virtual community [Maasberg et al. 2018][Osatuyi and Hughes 2018][Torres et al. 2018].

Due to this virtual live, it is natural to believe in information got in electronic vehicles more than we did before when the technology was not so well defined. Being even more common is the consumption, absorption and sharing of information without reliability check. Making us prone to propagate wrong information, such as fake news, hoaxes, etc.

However, more often than normal it is hard to differentiate what is a satire from what is an information intentionally fabricated to mislead us. So, in order to diminish false positives in automatic fake news, we carried an experiment born from an extensive fake news and sarcasm systematic literature review prior to this work, that consists of applying the techniques brought by that survey and building of two classical neural networks to be trained to differentiate true, sarcastic and fake Portuguese news.

2. The Current Experiment

The current experiment is an implementation of a fake news detection process we proposed in our prior works in this research area (Figure 1).

For the gather data step the experiment takes data from three different very well known Portuguese news sources, one for each possible class our model tries to learn, those being True News, Fake News and Sarcastic News. Respectively for each we have Folha de São Paulo, E-farsa prelabelled fake news and Sensacionalista sarcastic and satirical content. The news were filtered from a given time period in order to focus our analysis.

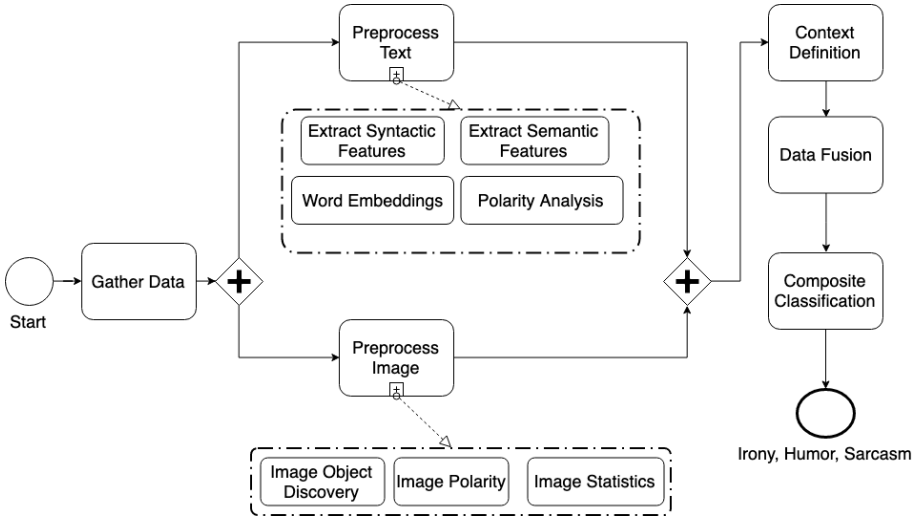

Figura 1. Fake News and Sarcasm Differentiation Process

The context chosen was the last Brazilian 2018 electoral run. We chose this period and political context due to the high volume of known Fake News being spread here in Brazil.

For the preprocess text step we applied a couple of Natural Language Techniques to extract the syntactic features, such as a dictionary counting capital letters, another counting special characters, a pronoun summary, tokenization of texts and sentences tokenization.

Then we proceed to the semantic features extraction which leverage on the previous step to check phrasal subjectivity and sentiment analysis.

Finally we used a pre-trained word embedding model known as Glove50 to encode our sentences into 20 dimension arrays, and in order to reduce the dimensionality of those we also used a recent proposed technique of t-sne, t-Distributed Stochastic Neighbor Embedding, a technique that implement dimensional reduction with better performance than the traditional Principal Component Analysis (PCA) technique.

After accuracy tests and literature review over the embedding matter, we tried a new approach of substituting this vectorization by sentence, to vectorization by document. So basically we used a prebuilt library(Word2Vec) to embed our documents into 50 dimension vectors, easing the learning process of our models, since they handle numerical data and not textual data.

For this experiment we ignored the Preprocess Image subprocess, as the main focus of this experiment is to detect fake news first by textual input through syntax and semantic, as we see the image processing as a future endeavor.

Then, for the context definition we fuse all the data preprocessed with the original metadata of each news and achieve the desired context of time and category before mentioned. Then pass to the next level of our process that is the current state of our research, the implementation of a composite classification, or what the literature identifies as an ensemble method.

In order to achieve the composite classifier we chose first to implement two neural

networks to understand the benefits our process has brought us. The chosen models were a simplistic feed forward neural network and a recurrent neural network. By choosing those specif networks we tried to understand a detection based on the pure analysis of the engineered features and the document word embedded arrays, and on the other hand, the second neural structure tries to take in consideration the context of document word embedded arrays in order to provide us better insights.

3. Construction of Baseline and Sanity Check

Since we are proposing a methodology, we want to establish a grounded truth by using the state of art techniques in the area, also, since we are tackling a classification problem we applied some sanity checks using unsupervised clustering algorithms just to guarantee that our data is also compliant to our ambitions, i.e. having the separation by classes we suspect it should have(True News, Sarcastic News and Fake News).

In order to have a baseline to compare our model to, we chose four classic models used by related works studied in prior surveys, those being KNN, Gaussian Naive Bayes, Random Forests and Support Vector Machine (SVM). Then we trained and scored those models against different instances of our dataset passing through the data fusion steps. With this we could see if the data fusion process was increasing the models' accuracy scores or not.

By monitoring the models evolution we were able to establish a baseline of accuracy scores in Table 1. Each column represents an accuracy score for each model in each row, from left to right we can follow the data fusion process from raw dataset, i.e. scraped data from news and numerical features such as length of text, then NLP enriched dataset, i.e. syntatic summary, semantic summary, special character counting, and sentiment analysis, then finally the document embedding step that encodes the document to 50 dimensions, i.e. Coordinates Dataset column and Fused Dataset column are respectively the scores for a dataset composed only by the 50 dimension coordinates and class, and the other one the final fused dataset.

We were able to see the significant increase of accuracy during each data fusion step, so proving that we were in the right direction, and also, establishing a baseline to be used as comparison against future composite approaches.

Models	Raw Dataset	NLP Dataset	Coordinates Dataset	Fused Dataset
Gaussian Naive Bayes	0,70	0,78	0,65	0,80
Support Vector Machine	0,66	0,90	0,66	0,90
Random Forest	0,71	0,63	0,81	0,64
K Nearest Neighbors	0,67	0,83	0,85	0,83

Tabela 1. Classic Models Baselines for each data fusion step of the detection process

For sanity check also, we ran Kmeans and DBScan unsupervised algorithms against unlabeled dataset in order to check that there is a intuitive separation of our documents into the three classes our model tries to classify, True News, Sarcastic News, and Fake News[Dhillon et al. 2003][Janssens et al. 2009]. And the results were what we expected, three well defined clusters.

4. Contributions

Through our efforts in this research project we believe that our main contributions are the detection process proposed, a prelabeled set of Portuguese News in the political context, and also the baseline models and neural networks trained with our enriched data.

As future contributions, we believe that our results will be a more robust model(the composite model), the preprocessed complete dataset of labeled News, and a new method of classifying Portuguese News, something not well explored nor defined nowadays, as most of the works in the area not only rely on social media topology analysis, but, also restrict themselves to only classify Fake News sources than classifying news' content.

5. Discussion

From the individual running of our chosen models we achieved great accuracy on both of around 98%(beating literature's current accuracy of averaging 80%), and also having low loss values of being 0.05 in average of both our models.

Although the results so far are satisfactory if compared against the literature, we are still experimenting other approaches such as convolutional neural networks that have word embedding and contextual layers. But, not yet done.

6. Future Works

For future works we intend to extend our efforts in finishing the composite classifier by also ensemble classical models' scores such as the Decision Tree, Random Forest and Support Vector Machine(SVM) as they were considered baselines.

Also, since the spread of fake news is a phenomena provoked, or at least aided by users' collaboration in between them and the author of the fake news, we want to modify our gather data step to be able to emulate or implement a collaborative system approach, as it would be more aligned to the reality.

Referências

Dhillon, I. S., Mallela, S., and Kumar, R. (2003). A divisive information-theoretic feature clustering algorithm for text classification. *Journal of machine learning research*, 3(Mar):1265–1287.

Janssens, F., Zhang, L., De Moor, B., and Glänzel, W. (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Information Processing & Management*, 45(6):683–702.

Maasberg, M., Ayaburi, E., Liu, C., and Au, Y. (2018). Exploring the propagation of fake cyber news: An experimental approach.

Osatuyi, B. and Hughes, J. (2018). A tale of two internet news platforms-real vs. fake: An elaboration likelihood model perspective. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Torres, R., Gerhart, N., and Negahban, A. (2018). Combating fake news: An investigation of information verification behaviors on social networking sites. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.