UNIVERSITÉ DE ROUEN

UFR SCIENCES ET TECHNIQUES

MASTER SCIENCE ET INGÉNIERIE DES DONNÉES

INSA - INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE ROUEN

LITIS - LABORATOIRE D'INFORMATIQUE, DE TRAITEMENT DE L'INFORMATION ET DES SYSTÈMES

# Deep learning

# Synthèse bibliographique

# "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection

12 décembre 2019

Authors :

Lucas ANQUETIL

El Hadji Brane SECK

Pierre LOPEZ

# Brief

# Introduction

The dissemination of fake news is not a new phenomenon, but has been present throughout the history of humanity. However, This phenomenon has taken on a new dimension in recent years, particularly with the advent of social networks, which make it possible to exchange information very quickly to a very large mass of people without verifying and regulating this information.

These fake news, which may come from one or more entities at different institutional levels (isolated person, media, government...), aim to misinform, often for a specific objective (destabilization of a government, search for financial advantages, election rigging...). These fake news are generally very difficult to detect a priori and can be created in such a way that they are as realistic as possible. One of the objectives would be to be able to automatically detect them as soon as possible in order to avoid their spread. Despite advances in automatic language processing, and the efforts made by researchers, scientific obstacles are still present.

One of the major problems was that there were very few dataset labelled to train models to detect fake news.

# 1. Dataset presentation

The article *Liar, Liar Pants on fire: A new Benchmark dataset for fake news Detection* imposes a dataset containing 12,836 short manually labelled statements from different sources (facebook, twitter, political debates, etc.).

Each statement is associated with a series of information (which will be called metadata later on). They are as follows : the author of the declaration, its political party, the subject, the context, the place. Then each declaration is also associated with a label indicating the veracity of the text.

The labels are as follows:

- Pants on fire
- False
- Barely true
- Half true
- Mostly true
- True

**Statement**: *"Newly Elected Republican Senators Sign Pledge to Eliminate Food Stamp Program in 2015."*
**Speaker**: Facebook posts
**Context**: social media posting
**Label**: Pants on Fire

**Statement**: *"Under the health care law, everybody will have lower rates, better quality care and better access."*
**Speaker**: Nancy Pelosi
**Context**: on 'Meet the Press'
**Label**: False

*Figure 1 : Two example of statement, information and labels associated*

The proposed dataset is already split into a training, validation and test set. Each class is well balanced and this in each of the set.

| Dataset Statistics | |
|---|---|
| Training set size | 10,269 |
| Validation set size | 1,284 |
| Testing set size | 1,283 |
| Avg. statement length (tokens) | 17.9 |

*Figure 2 : Statistics of the LIAR dataset*

The authors of the article then question the possibility of using a deep learning model to detect fake news using this dataset. Their contribution was as follows : the dataset is composed of 6 labels so the goal is to make models performing multi-class classification on text.

Initially, the authors tested classifiers, only on statements (not including metadata), such as SVMs, logistic regression classifier, etc. Then, they introduced the use of deep models. They first made a pre-trained word embedding from Google News on statements and then used CNNs and B-LSTMs for example.

# 2. Innovations

This article introduces the LIAR dataset which, in addition to the declarations, provides us with a series of information about this declaration as mentioned above. Until now, few dataset included this type of information. The fact that we have about 12.8K statements with this information gives us the opportunity to test deep learning methods.

A second feature of this dataset is that the labels are not just binary (either True or False). Indeed, it has 6 more specific classes.

Consequently, the authors were able to test deep learning models on this dataset by including metadata and performing a multi-class classification of text task. All this to see if NLP methods could be improved by providing additional information (metadata).

# 3. Proposed model

In the article, the authors propose an architecture that allows metadata to be included to improve classification.

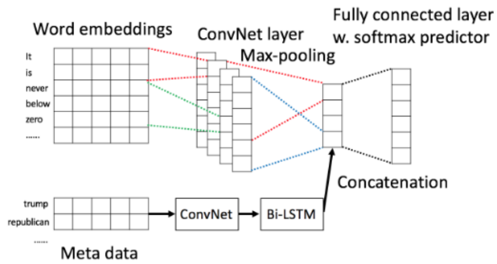


*Figure 3 : Architecture of the proposed model of the article*

They use ConvNet and a B-LSTM on metatdata embeddings. The output of this will be concatenate to the output of the conv block (on the statements) and the prediction vector will be obtained through a Dense with softmax activation.

| Models | Valid. | Test |
|---|---|---|
| Majority | 0.204 | 0.208 |
| SVMs | 0.258 | 0.255 |
| Logistic Regress0ion | 0.257 | 0.247 |
| Bi-LSTMs | 0.223 | 0.233 |
| CNNs | 0.260 | 0.270 |
| Hybrid CNNs | | |
| Text + Subject | 0.263 | 0.235 |
| Text + Speaker | **0.277** | 0.248 |
| Text + Job | 0.270 | 0.258 |
| Text + State | 0.246 | 0.256 |
| Text + Party | 0.259 | 0.248 |
| Text + Context | 0.251 | 0.243 |
| Text + History | 0.246 | 0.241 |
| Text + All | 0.247 | **0.274** |

*Figure 3 : Evaluation results on the LIAR dataset.*
*The top section : statement-only models.*

We observe that the use of the proposed architecture does not significantly improve performance. Consequently, questions appear about the daatset, the architecture and the approach of the problem. This is what we will see in the next part.

# 4. Criticisms and limits

First of all, having 6 possible labels seemed like a good thing to be more precise and not only to predict TRUE or FALSE, but we notice that the annotations made manually by humans can differ according to perception. We want to be thinner but it also brings doubt for our classification task.

Then, one of the defects of the article comes from the fact that the authors do not go into enough detail about how they managed to classify with SVM for example. Indeed, statements are variable in length so we are entitled to wonder how they managed this to apply a classification with an SVM. Precisely it must be the introduction of an LSTM that manages well the cases where the size of the input sequences is variable, and in the article there is no detail on this point.

In addition to this, the details of their proposed architecture are not submitted. Indeed, once again, they apply a Conv to the word embedding of statements but how do they manage word vectors of varying size only with a Conv ?

Moreover, when introducing metadata, the article states the following sentence : "We randomly initialize a matrix of embedding vectors to encode the metadata embeddings. We use a convolutional layer to capture the dependency between the metadata vector(s). Then,a standard max-pooling operation is performed on the latent space, followed by a bi-directional LSTM layer.". Here, we can understand the use of a Conv but we can ask ourselves the question: what does the B-LSTM do and what is it used for metadata?

Finally, the main problem with this article is the approach taken by the authors. Indeed, it presents the problem as a multi-class classification task. However, we have in the LIAR dataset, ordered classes, scaled from lie to truth. When performing classification, predicting the false and mostly-true class on a statement (ground truth : True) represents the same error while in the theory mostly-ture prediction is much read close to the ground-truth. Therefore, the approach is bad and in the end here it is a task of regression that must be done. We can take inspiration from the technique used in this paper *Rank-consistent Ordinal Regression for Neural Networks* to do this. The authors describe a technique for performing ordinal regression, i. e. predicting labels on an ordinal scale using a ranking rule.

# Conclusion

The detection of fake news is a complex problem. Indeed, it is necessary to have labelled data while trying to have an explanatory conetxte. The LIAR dataset presented in this article tries to bring additional information to the text in order to better detect fake news but we notice the approach used was not the box on this dataset. Thus the results and the conclusion of the article that metatdat can improve the performance of a model cannot be maintained because the results are biased.

We can also ask the following question: can we really have good performance by using networks on the text and on metatdata limited to those of the LIAR dataset ? To really improve fake news detection performance, it is probably necessary to include an external information flow and more context to the stamements.

In addition, this kind of dataset like the LIAR dataset is labeled by humans. And it should be remembered that there is no clear limit between a fake news and a factual article. Rather, it is a continuum on which a cursor must be placed. A political opinion or an advertisement can be considered as fake news. Any militant or political statement can play with the truth.

Then just apply NLP method on statements does not allow to detect fake news. Then the the metadata do not necessarily have any meaning.