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

Report

Fake news detection

---

3 novembre 2019

Lucas Anquetil
Pierre LOPEZ
El Hadji Brane SECK

INSA Rouen - Département ASI

Table des matières

Introduction

The dissemination of false news is not a new phenomenon, but has been present throughout the history of humanity: historians have found traces of it since antiquity. 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...). 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 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.

1. Presentation of the article

1.1 Dataset presentation

In this article, we find a dataset labeled for fake news detection. Nowadays, this identification is necessary but remains a complicated problem to solve. This new dataset will allow us to apply deep learning methods to detect fake news automatically. The data were collected in different contexts such as political debates, facebook and twitter posts, etc.

Therefore, this dataset contains shorts statements and information on the subject, context/venue, speaker, state, party. Each short statement is labelled by 6 possible fine-grained classes for the truthfulness ratings :

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

In this dataset, labels are well-balanced. This dataset contains 12,836 examples.

Dataset Statistics	
Training set size	10,269
Validation set size	1,284
Testing set size	1,283
Avg. statement length (tokens)	17.9
Top-3 Speaker Affiliations	
Democrats	4,150
Republicans	5,687
None (e.g., FB posts)	2,185

Figure 1 : The LIAR dataset statistics

An example of a sample :

Statement: *"The last quarter, it was just announced, our gross domestic product was below zero. Who ever heard of this? Its never below zero."*
Speaker: Donald Trump
Context: presidential announcement speech
Label: Pants on Fire
Justification: According to Bureau of Economic Analysis and National Bureau of Economic Research, the growth in the gross domestic product has been below zero 42 times over 68 years. Thats a lot more than "never." We rate his claim Pants on Fire!

Figure 2 : Sample on the LIAR dataset

1.2 Method proposed in the article

In the article, a proposal for a deep neural network to solve this multiclass text classification problem is made.

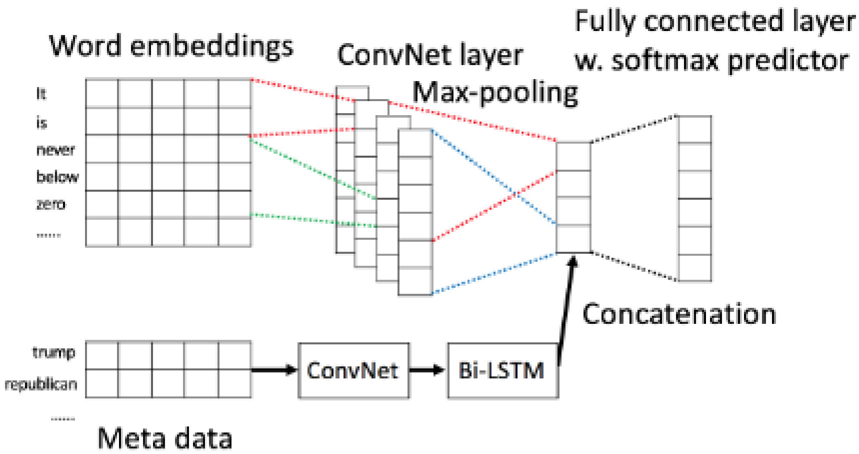

Figure 3 : Proposed architecture of the article (called Hybrid CNNs)

The details of the implemetation are in the article. In the following table, you will find different result with basics methods and with the proposed method in the article.

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 + All	0.247	**0.274**

Figure 4 : Results on the LIAR dataset

2. Our realizations

2.1 First approach

Our first objective was to reproduce the model of the article as well as simple models without including metadata.

We used pre-trained 100-dimensional word2vec embeddings from Glove. Then we tested with different pre-trained Glove but this had no influence on the performance.

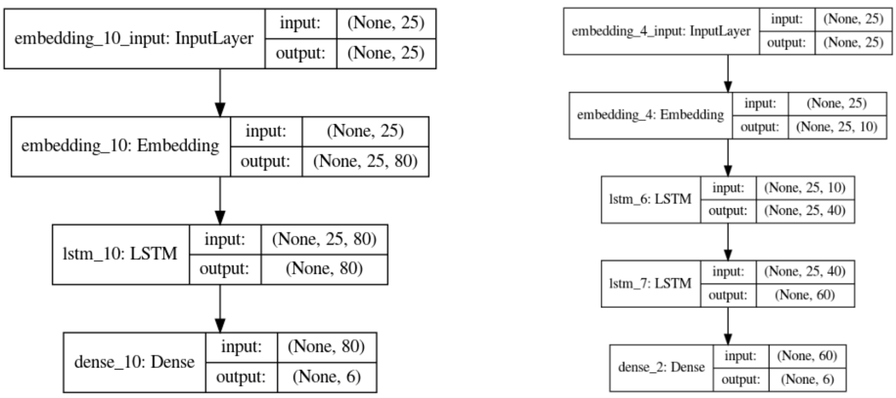

Figure 5 : Two examples our proposed architectures without metadata included

These simple models are found in the github (idx_model from 14 to 19). Inside, we find different experiments with parameter changes. We also find the results and the behaviour.

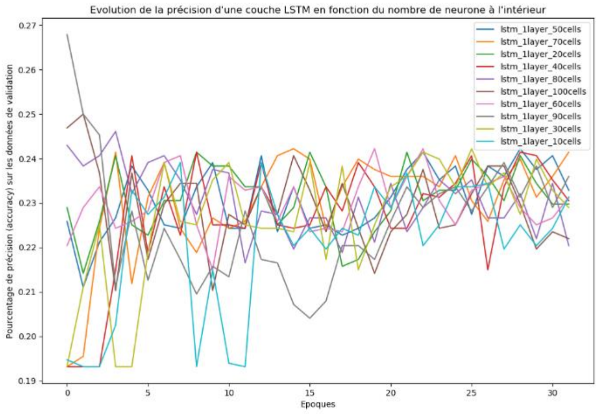

Figure 6 : Behaviour of LSTM based models with different layers size during learning

It can be observed that on the basis of validation, there is no convergence during learning. Therefore, performance is not good and an apprentice on statements may not be enough.

The goal will be to integrate the metadata. We will see the performances in the next section.

2.2 Second approach

In this section, we will present our more complex models by including metadata (idx_model from 1 to 13).

We decided to take inspiration from this architecture to implement our own architectures. By using LSTMs, CNNs, varying the size of the layers and also by using different dataset for embedding.

- Input : word embedding of statements (+ metadata)
- Loss : categorical_entropy
- Output : softmax predictor

Evaluate with :

- metric : categorical_accuracy

For metadata, we perform a pre-treatment in order to pass the speakers, venue, context, etc. in scalar. Thus, we will be able to concatenate the metadata in our model to see if they provide better performance.

We use RNNs or CNNs and sometimes both.

Here are some examples below:

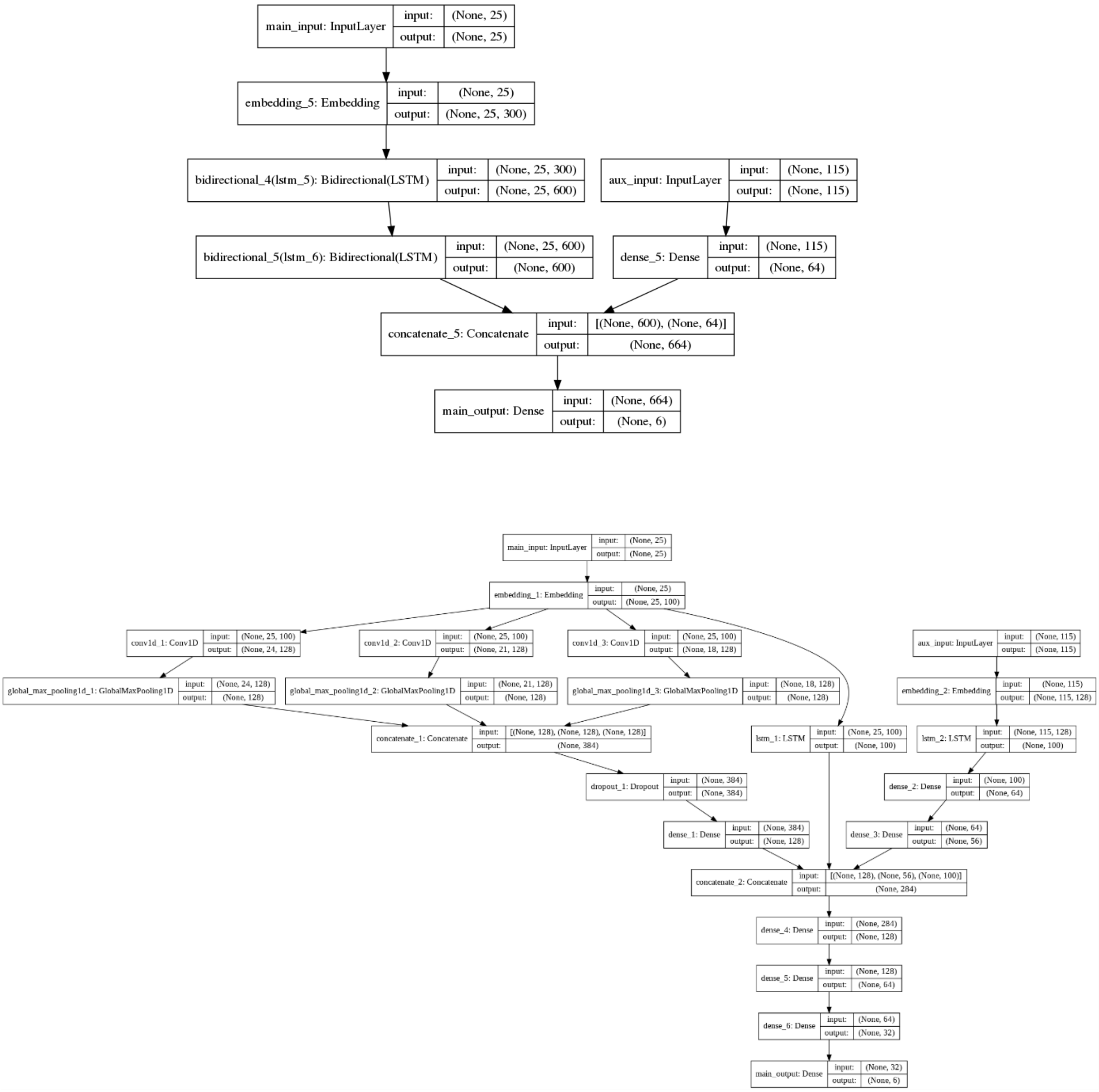

Figure 7 : Two examples our proposed architectures with metadata included

And below, we find an example of the behaviour of these models.

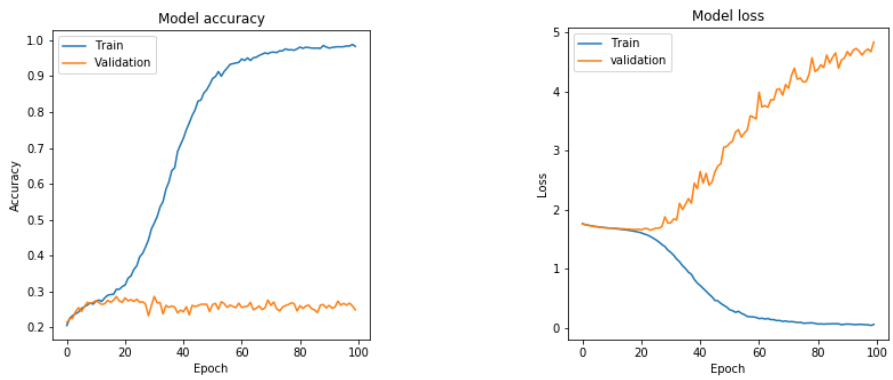

Figure 7 : Accuracy and loss evolution during training

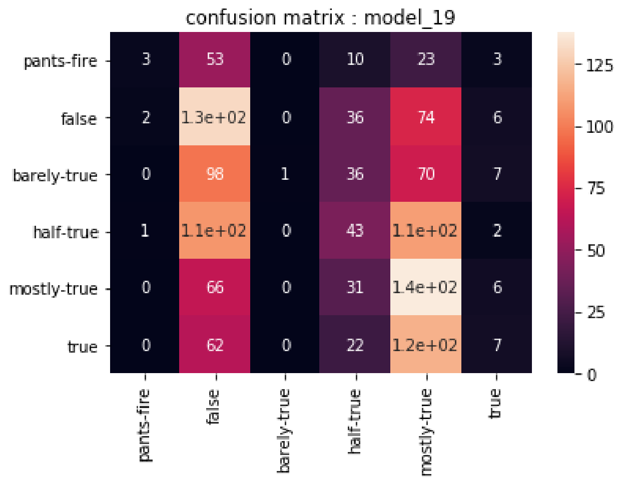

Figure 8 : Confusion matrix for the test dataset

All details and results can be found under the github.

But the behaviour of all models is similar and we notice that we are falling into overlearning. In addition, we achieve the same performance of the article. Moreover, seeing confusion matrix, class predicted are not well balanced.

Results with categorical labels (categorical accuracy)			
idx-model	meta-data	description	Test
3	yes	CNN (statement) + Dense (meta)	0.288
13	yes	2xBLSTM (statement) + Dense (meta)	0.277
14	no	BLSTM(60) (statement)	0.262
...	...	...	...
19	no	CNN (statement)	0.255
8	yes	CNN+LSTM(statement)+LSTM(meta)	0.245
...	...	...	...
18	no	2xBLSTM(statement)	0.249
15	no	GRU(statement)	0.242

Figure 9 : Table showing few results on the test performance for few models

We note that the complexity of the model does not necessarily allow for improved performance and that the addition of metadata does not necessarily provide additional information and therefore no better performance.

2.3 Binarised approach

Seeing the behavior of our networks on the 6 classes, we decided to modify our labels and just put True and False.

Then we decided to see the results for the models previously implemented.

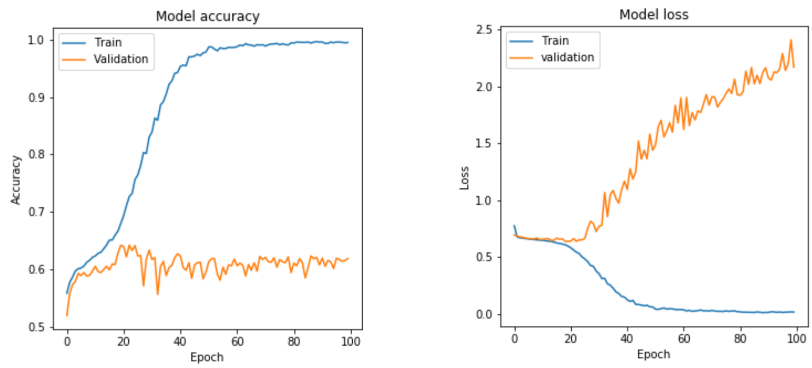

Figure 10 : Accuracy and loss evolution during training

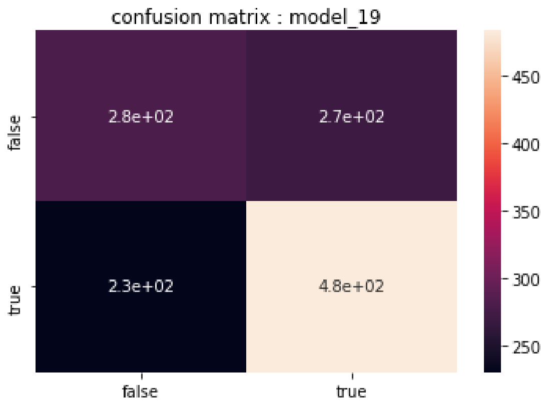

Figure 11 : Confusion matrix for the test dataset

Results with categorical labels (categorical accuracy)			
idx-model	meta-data	description	Test
3	yes	CNN (statement) + Dense (meta)	0.648
13	yes	2xBLSTM (statement) + Dense (meta)	0.635
14	no	BLSTM(60) (statement)	0.626
...	...	...	...
19	no	CNN (statement)	0.621
8	yes	CNN+LSTM(statement)+LSTM(meta)	0.616
...	...	...	...
18	no	2xBLSTM(statement)	0.615
15	no	GRU(statement)	0.61

We see that performance remains poor. And then we can ask ourselves questions about the dataset and metadata. That's what we're going to see in the next game.

3. Perspectives and limits

Fake news detection is a complex problem. Having tested different RNNs or CNNs architectures or both on text only does not allow you to determine fake news. Indeed, to determine this we have to look after external elements. However, the metadata of our dataset are not explicit enough to improve performance and so we have seen by including them that they do not necessarily have any influences.

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 metadata do not necessarily have any meaning.