Detecting Fake news using NLP techniques
Machine Learning for Natural Language Processing 2020

Lucas ALLARD
ENSAE
lucas.allard@ensae.fr

Wissal EL ACHOURI
ENSAE
wissal.elachouri@ensae.fr

Abstract

[1] [2] Our work aims to classify fake news. Our dataset contains more than 20 000 news and among them, half is real news and the other half is fake news. We have processed the texts of the comments using three different embedding vectorizers (TF-IDF, Word2Vec and BERT). For the classification of fake news, in the first time, we have used four classic models (logistic regression, SVM, RF and naive Bayes) using TF-IDF embedding. In the second part, we have used LSTM and BiLSTM models using Word2Vec embedding. And then, we have used the BERT transformer in the last section.

1 Problem Framing

The rise of Fake news is indeniable. Whether it's through traditionnal outlets or social media channels, this growing issue has serious consequences that results in the public being exposed to false narratives that could lead potentially to violence and health hasards. Our aim is to address this problem through a Natural Language Process approach by focusing only on the content being reported. In the first place, we start by doing some data exploration to exhibit the common themes present in fake articles mainly via a Latent Dirichlet analysis then we propose a bunch of methods to detect the realibility of a news article ranging from simple classifiers to a more complex method : Bert Transfomer. The data set that we used consists of about 20 000 articles containing a combination of real and fake articles. The attributes of the dataset are :
- id: which is a unique identifier to the article
- title: the headline of the article
- author: the author of the article
- text : the content of the article
- label: a label that indicates if the article is potentially unreliable 1: or 0: reliable

2 Experiments Protocol

Our overall approach to text classification was starting with a simple baseline model and gradually adding more sophisticated features to see if and how the results improve. Each method is described concisely below.

2.1 LDA

We built an insightful topic model based on the Latent Dirichlet Allocation (LDA) (1) algorithm.

2.2 Traditional Approaches

We used term frequency-inverse document frequency (tf-idf) (2) as our main embedding and combined it with the following machine learning algorithms:

- Logistic Regression
- Support Vector Machine
- Naive Bayes
- Random Forest

2.3 Word Embedding + LSTM

The model uses a pre-trained 100-dimensional Word2Vec (3) embeddings, which are not updated during training, given that the training set is small compared to the magnitude of the corpora used to train Word2Vec.

2.4 BERT Transformer

We used a pre-trained BERT (4), one of the most popular transformer models, and fine-tune it on fake news detection to achieve state of the art results

[1] https://github.com/siwills/NLP_Fake_news

[2] https://colab.research.google.com/drive/1gMpyG-36L_kY4L0u8GtiGYIE9XZ83DoK?usp=sharing

3 Results

The performance of each model employed in this project is summarized in table: There are few points worth noting about these results. TF-IDF and machine learning methods yielded good results although they are quite simple campred to LSTM and BERT.

Method	Accuracy	F1-Score
Logistic Regression	0.94	0.95
SVM	0.96	0.96
Naive Bayes	0.86	0.84
Random Forest	0.92	0.92
Word2Vec + LSTM	0.94	0.93
Word2Vec + BI LSTM	0.95	0.96
BERT	0.99	0.99

4 Discussion/Conclusion

In future work and to improve the quality of our classification, we would like to incorporate different features beyond the text. For fake news detection, we can add as features the source of the news, the topic, the publishing plateform. We would have ideally explored the weights assigned to the used models to better understand how a given corpus is weighted.

References

[1] Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, Liang Zhao. *Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey*. 2017.

[2] Cedric De Booma, Steven Van Canneyta, Thomas Demeestera, Bart Dhoedta. *Representation learning for very short texts using weighted word embedding aggregation*. 2016.

[3] Tomas Mikolov, Ilya Sutskever, Kai Chen al. *Distributed Representations of Words and Phrases and their Compositionality*. Google Inc, Mountain View, 2013.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2018.