Fake News Detection Using Machine Learning

Author: Simon Lorent
Supervisor: Aswhin Ittoo

September 10, 2019

University Of Liège

Introduction

Fake news used to influence elections

2016 US presidential elections (Allcott et al.)[1]

- 62% of US citizens get their news for social medias[2]
- Fake news had more share on Facebook than mainstream news[3].

1. Expliquer l'inpact des fake news sur les elections
2. La suite dfinit ce que sont les fake news

Definition

Definition

Fake news is a news article that is intentionally and verifiable false[4]

1. Dire que l'on peut caracteriser les fake news de plusieurs faon: le contenu et le context

Fake News Characterisation I

News content features:

- **Source**: Where does the news come from, who wrote it, is this source reliable or not.
- **Headline**: Short summary of the news content that try to attract the reader.
- **Body Text**: The actual text content of the news.
- **Image/Video**: Usualy, textual information is agremented with visual information such as images, videos or audio.

Fake News Characterisation II

Fake News Characterisation III

Different kind of models:

- **Expert-oriented**: relies on experts, such as journalists or scientists, to assess the news content.
- **Crowdsourcing-oriented**: relies on the wisdom of crowd that says that if a sufficiently large number of persons say that something is false or true then it should be.
- **Computational-oriented**: relies on automatic fact checking, that could be based on external resources such as DBpedia.

Methodology

Methodology I

Goal:

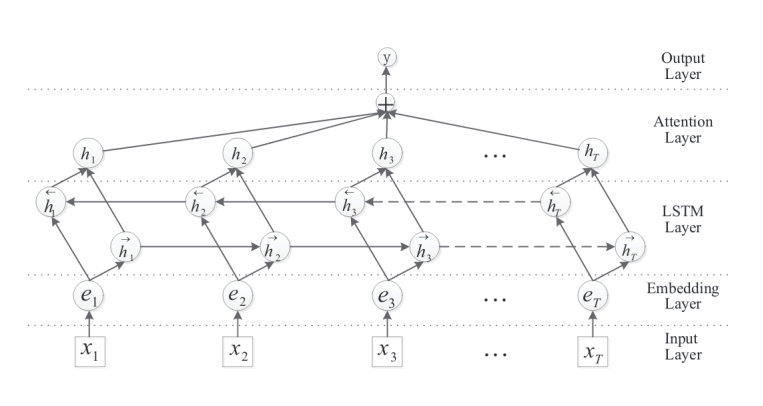
- Comparing the performances of ”traditional” machine learning techniques and deep learning techniques
- Comparing the performances of these techniques on two differents datasets.

Methodology II

Models used:

- Naïve-Bayes
- SVM
- Decision Tree
- Ridge Classifier
- LSTM
- Attention Mechanism[5]

Attention Mechanism I



Word Embedding

- One-hot encoding
- Word2Vec[6]

- $e_i = W_e * x_i$
- $e_i =$ dictionary lookup

W_e is a tunnable parameters and word2vec are pretrainedOne-hot vector are made from the training set by building a dictionary

Attention Mechanism

$$H = [h_1, h_2, ..., h_T] \tag{1}$$

Where T is the sequence length. Then we define

$$M = \tanh(H) \tag{2}$$

$$\alpha = softmax(w^T M) \tag{3}$$

$$r = H\alpha^T \tag{4}$$

Finally, we compute $h^* = \tanh(r)$.

Results

Machine learning

	fake	reliable	accuracy	macro avg	weighted avg
f1-score	0.514399	0.679764	0.614049	0.597082	0.607588
precision	0.570485	0.638376	0.614049	0.604430	0.608744
recall	0.468354	0.726891	0.614049	0.597623	0.614049
support	1106.000000	1428.000000	0.614049	2534.000000	2534.000000

(a) Raw results for Linear SVM

	fake	reliable	accuracy	macro avg	weighted avg
f1-score	0.412107	0.698507	0.601421	0.555307	0.573504
precision	0.578431	0.608741	0.601421	0.593586	0.595512
recall	0.320072	0.819328	0.601421	0.569700	0.601421
support	1106.000000	1428.000000	0.601421	2534.000000	2534.000000

(b) Raw results for Naïve-Bayes

	fake	reliable	accuracy	macro avg	weighted avg
f1-score	0.496366	0.691279	0.617206	0.593822	0.606207
precision	0.582927	0.633606	0.617206	0.608266	0.611486
recall	0.432188	0.760504	0.617206	0.596346	0.617206
support	1106.000000	1428.000000	0.617206	2534.000000	2534.000000

(c) Raw results for Ridge Classifer.

	fake	reliable	accuracy	macro avg	weighted avg
f1-score	0.479354	0.591549	0.542226	0.535451	0.542580
precision	0.475936	0.594901	0.542226	0.535418	0.542977
recall	0.482821	0.588235	0.542226	0.535528	0.542226
support	1106.000000	1428.000000	0.542226	2534.000000	2534.000000

(d) Raw results for Decision Tree

Table 4.1: Raw results on **Liar-Liar Corpus**.

	fake	reliable	accuracy	macro avg	weighted avg
f1-score	0.894700	0.965364	0.947874	0.930032	0.947620
precision	0.907861	0.960783	0.947874	0.934322	0.947494
recall	0.881916	0.969989	0.947874	0.925952	0.947874
support	17496.000000	52181.000000	0.947874	69677.000000	69677.000000

(a) Raw results for Linear SVM on **Fake News Corpus**

	fake	reliable	accuracy	macro avg	weighted avg
f1-score	0.674458	0.905634	0.853682	0.790046	0.847585
precision	0.764127	0.875841	0.853682	0.819984	0.847790
recall	0.603624	0.937525	0.853682	0.770574	0.853682
support	17496.000000	52181.000000	0.853682	69677.000000	69677.000000

(b) Raw results for Naïve-Bayes on **Fake News Corpus**

	fake	reliable	accuracy	macro avg	weighted avg
f1-score	0.874220	0.960438	0.939808	0.917329	0.938788
precision	0.919674	0.945736	0.939808	0.932705	0.939192
recall	0.833048	0.975604	0.939808	0.904326	0.939808
support	17496.000000	52181.000000	0.939808	69677.000000	69677.000000

(c) Raw results for Ridge Classifier on **Fake News Corpus**

	fake	reliable	accuracy	macro avg	weighted avg
f1-score	0.791687	0.929799	0.894987	0.860743	0.895119
precision	0.788700	0.930987	0.894987	0.859844	0.895258
recall	0.794696	0.928614	0.894987	0.861655	0.894987
support	17496.000000	52181.000000	0.894987	69677.000000	69677.000000

(d) Raw results for Decision Tree on **Fake News Corpus**

Table 4.2: Results on **Fake News Corpus** without using SMOTE.

Machine learning III

	fake	reliable	accuracy	macro avg	weighted avg
f1-score	0.891373	0.962340	0.94407	0.926857	0.944520
precision	0.869960	0.970623	0.94407	0.920291	0.945346
recall	0.913866	0.954198	0.94407	0.934032	0.944070
support	17496.000000	52181.000000	0.94407	69677.000000	69677.000000

(a) Raw results of linear svm on **Fake News Corpus** when training using SMOTE

	fake	reliable	accuracy	macro avg	weighted avg
f1-score	0.714816	0.873538	0.824777	0.794177	0.833683
precision	0.604424	0.950521	0.824777	0.777472	0.863615
recall	0.874543	0.808091	0.824777	0.841317	0.824777
support	17496.000000	52181.000000	0.824777	69677.000000	69677.000000

(b) Raw results of Naïve-Bayes on **Fake News Corpus** when training using SMOTE

	fake	reliable	accuracy	macro avg	weighted avg
f1-score	0.877755	0.956129	0.935431	0.916942	0.936449
precision	0.836588	0.973317	0.935431	0.904953	0.938984
recall	0.923182	0.939537	0.935431	0.931360	0.935431
support	17496.000000	52181.000000	0.935431	69677.000000	69677.000000

(c) Raw results of Ridge Classifier on **Fake News Corpus** when training using SMOTE

	fake	reliable	accuracy	macro avg	weighted avg
f1-score	0.787226	0.921178	0.884969	0.854202	0.887542
precision	0.734992	0.946085	0.884969	0.840539	0.893079
recall	0.847451	0.897549	0.884969	0.872500	0.884969
support	17496.000000	52181.000000	0.884969	69677.000000	69677.000000

(d) Raw results of Decision tree on **Fake News Corpus** when training using SMOTE

Table 4.3: Results on **Fake News Corpus** when training with SMOTE.

Bibliography I

- Hunt Allcott and Matthew Gentzkow.
  Social media and fake news in the 2016 election.
  In *Journal of Economic Perspective*, volume 31, 2017.
- Jeffrey Gottfried and Elisa Shearer.
  ***News Use Across Social Medial Platforms 2016*.**
  Pew Research Center, 2016.
- Craig Silverman and Lawrence Alexander.
  How teens in the balkans are duping trump supporters with fake news.
  Buzzfeed News, 3, 2016.

Bibliography II

- Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu.
  Fake news detection on social media: A data mining perspective.
  ACM SIGKDD Explorations Newsletter, 19(1):22–36, 2017.
- Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu.
  Attention-based bidirectional long short-term memory networks for relation classification.
  In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short*

Bibliography III

Papers), pages 207–212, Berlin, Germany, August 2016. Association for Computational Linguistics.


Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean.

Efficient estimation of word representations in vector space.