Fake News Detection

Kabir Kakkar[#1], Manoj Kumar R[#2], Jayant Singh Rathor[#3]

ABSTRACT *Fake news is news that is produced purposely to confuse viewers. It is a kind of propaganda in the nature of real news that is written. Via conventional news sources and social media, fake news is distributed. Fake news has long been a concern. The quality of information on the Internet, especially on social media, is an increasingly important issue, but the ability to detect, analyse and correct such evidence, or so-called "false news," present on these sites, is hindered by web-scale data. We present in this paper a tool for the analysis of "false news". To predict whether a given article will be classified as True or False, this approach utilises the Multinomial Naive Bayes classification model and the construction of the famous LSTM neural network to get the appropriate fit for the model used to classify our article. Through applying many methods that are discussed in the article, the results can be improved. The findings obtained show that the issue of false news identification can be solved using machine learning techniques.*

1. Introduction

Fake news is news that is created to mislead viewers purposefully. In the type of real news, it is a kind of misinformation that is written. Fake news is being spread via traditional news outlets and social media. For a long time, fake news has been an issue. The dissemination of false news has intensified with the advent of social media, and it has become hard to distinguish between real news and fake news. As it manipulates public views, the dissemination of false news is a matter of concern. During the 2016 American Presidential elections, more than 1 million tweets were estimated to be tied to "Pizzagate" fake news by the end of the election.

The widespread distribution of false news may have a significant detrimental effect on people and culture. As it is far more commonly circulated on social media than most mainstream authentic news, it has taken down the credibility of the news ecosystem. It is one of the main topics that can alter attitudes and affect decisions and disrupt the way people respond to real news. It has political impact, may foster mistrust of legitimate media institutions which may affect capital markets and hurt the credibility of a person. A study reported during the 2016 American presidential election that many young men and teenagers in Veles ran hundreds of websites that released many fake viral reports that endorsed Trump. Such fake news infected numerous persons who impacted the outcome of the referendum. This is only a small instance of how persons can be affected by the dissemination of false news.

In attempts to avoid the dissemination of fake news, several organisations have come forward. Just ex. Eg. Artificial Intelligence is used by the Google app to pick storeys and combat fake news. Using machine learning and NLP techniques, we strive to create a model to decide if a news is false or accurate.

Fake news is a major threat as it can cause fear in the population easily. As was seen in the US Presidential Polls, it can also impact major world affairs. It is difficult to validate news using conventional fact checkers and vetting, with the influx of news coming from web content generators, as well as diverse formats and genres. We have a statistical instrument to resolve this issue of rapid and reliable labelling of news as false or authentic.
The major objectives of this project are:

1. Taking the help of linguistic cues to develop a machine learning based model for accurately determining whether the given news is fake or authentic.
2. To get high accuracy to determine a news is fake or true.

2. Literature Review

In the research paper "Syntactic Stylometry for Deception Detection" by
Feng, Song and Banerjee, Ritwik and Choi, Yejin, rely in their language on the language habits followed by deceptors. Many liars strategically manipulate their vocabulary to prevent being detected, despite their effort to regulate what they speak, "leakage" language happens in some linguistic elements that are difficult to manually track, such as pronoun frequencies and patterns, conjunction, and use of negative emotion terms. The research focuses on
user reviews and essays, but is equally applicable in fake news detection as
well, where the author tries to deceive the readers. For detecting these linguistic patterns, the research uses shallow and deep syntax analysis which use POS(parts-of-speech) tags and Probabilistic Context Free Grammars(PCFG) Respectively.

In their research paper "Automatic deception detection for detecting fake news" by Conroy, Niall J. and Rubin, Victoria L. and Chen, Yimin, they discussed two methods for Fake News detection. The first is a linguistic approach, which examines the different syntactic and semantic properties that are effective in detecting deceit. It uses deep syntax analysis using Stanford parser with context-free grammar generation. The foundation of the semantic interpretation given in this study is that when writing, the author may use inconsistencies and withhold evidence, whereas a writer of a factual review is more likely than other truthful reviewers to make identical statements regarding aspects of the product. The path to the network depends on querying current information networks or organised data that is publicly accessible, such as DBpedia ontology or Google Relation Extraction Corpus (GREC). It also takes into consideration that the authentication of the identity of the person writing an article is paramount to the principle of trust on social media.

In their research paper by Veronica Perez-Rosas , Bennett Kleinberg , Alexandra Lefevre Rada Mihalcea1, the emphasis was on automatically detecting bogus material in the online news. For the task of false news identification, they implemented two innovative databases, spanning seven distinct news domains. By removing multiple linguistic features such as n-grams,

punctuations, psycholinguistic features, readability, syntax etc., they create fake news identification models. To create these functions, they used LIWC. Their best models have attained accuracy equal to the human capacity to detect false news.

In their research paper "Evaluating Machine Learning Algorithms for Fake News Detection" by Shlok Gilda, he explored various Natural Language Processing techniques for the detection of fake news detection. He put in algorithms like TF-IDF of bi-grams and probabilistic context free grammar (PCFG) detection on a figure of data corpus which constituted 11,000 articles. He tested the dataset on multiple classification algorithms like Gradient Boosting, Bounded Decision Trees, SVM,Random Forests and Stochastic Gradient Descent. They discovered that the Stochastic Gradient Descent model came up with an accuracy of 77.2%.

The wide spread of fake news can have a huge negative impact on individuals and society as a whole. We developed a model which accurately determines whether an article is fake or not using machine learning and NLP techniques.

Datasets

We have used Fake News Dataset from Kaggle and used required attributes of these datasets to train our model.
The dataset has about 20,800 data points. But there are 20,242 data points after preprocessing the data, that is, eliminating duplicates and deleting data points that are NULL.

3. Methodology

The approach proposed for this project is:
1. Data Preprocessing
2. Generating News Feature Vector
3. Classification

Data Preprocessing:

Data preprocessing is done to convert data into the required format of raw data. Data pre-processing can be carried out using different techniques such as data washing, data elimination, integration of data, etc. The datasets are obtained in this project from numerous tools which have different formats and attributes. The data may then be duplicated and could contain certain attributes that are not useful. So with the appropriate attributes that are used to train our model, we convert the data to our desired format. The corpus is converted into tokens by splitting it into tokens. Since there would be certain words which would contribute relatively less we would remove those stop words as well. Then we used a porter stemmer for stemming.

Porter Stemming: The Porter stemming algorithm (or 'Porter stemmer') is a tool to extract from english words, the commoner morphological and inflexional terminations. Its primary usage is a faction of the meaning of the normalisation process.

Generating News Feature Vector:

Converting the news stories into a news vector including the essential features needed to assess the meaning of the news is the most important aspect of detecting whether or not a given news is false.

There are many ways for this function vector to be created. To decide which procedure gives the best precision, we tried various methods for the same. Some of the techniques are as follows:

Bag of Words

"Bag of Words" is a way of expressing text in a manner that can be quickly interpreted by machine learning algorithms. One of the methods of removing characteristics from text is BoW. There are primarily 2 items involved with this form of text representation:

1. Dictionary of familiar words
2. Part of the presence of familiar words.

In BoW representation, the order of words or structure of the sentence is not considered, it is only distressed with whether the term is commenced in the document or not. The BoW is carried out as follows:
Collect all the unique words in all the documents.
Now, represent the documents in a vector of all the unique terms by counting the number of times each term is present in that document.

TF-IDF:

TF-IDF stands for term frequency-inverse document frequency.TF-IDF is a method used to represent text in a format which can be easily processed by machine learning algorithms. It is a numerical statistic that shows how important a word is in a document in a word corpus. The importance of a word is proportional to the number of times the word appears in the document but inversely proportional to the number of times the word appears in the corpus. The tf-idf weight is composed of 2 terms:

1. **TF (Term Frequency):** Term frequency is specified in the document as the frequency of a phrase. TF is measured as: TF(w) = (In the text, the number of times the word 'w' appears) / (In the document, the total number of words).

2. **IDF (Inverse Document Frequency)**: This analyses in the paper how relevant a term is. Words such as and, and of, for example, a couple of occasions, but they are less significant. Thus less weights are given to most repetitive terms and more weights are given to less regular terms. The IDF is measured as:

   $IDF(w) = \log_e(\text{Total number of documents} / \text{Number of documents containing the word 'w'})$.

By computing TF*IDF values, the TF-IDF weight is assigned to each term.

We measure the tf-idf values of the bigrams and represent the tf-idf vector of such bigrams in order to produce the news vector. We choose bigrams over unigrams, since the context is given.

N-grams: The n-gram from a given text is a contiguous sequence of n words. Ann-gram size 1 is referred to as a "unigram"; a "bigram" is size 2, a "trigram" is size 3, and so on. A model will store more background with greater n.

7.3 Classification*:* After creating the vector for the news function, we now define the vector to be false or true. For the purpose of classification, we plan to use the following classification algorithms:

Multinomial-Naive Bayes*:* A supervised learning algorithm that is used for classification is the Multinomial-Naive Bayes. It is based on the theorem of Bayes, assuming that characteristics are independent of each other. It measures the likelihood of each class, choosing the class with highest probability as the output. Unlike Bayes, Multinomial distribution is used and for text-classification is highly efficient.

LSTM :

LSTM has a sequential model which is illustrated with a green box. if unfolded the architecture becomes as below:

The distinction between RNN and LSTM is that it has additional signal information, usually referred to as "cell memory," which is supplied from one point to the next. Using the gate mechanism, LSTM is designed to solve the issue of vanishing gradient.

LSTM Modeling

- Vectorize corpus text, by turning each text into either a sequence of integers or into a vector, using one hot encoding.
- Set the max number of words in each complaint at 25.
- Truncate and pad the input sequences so that they are all in the same length for modeling.
- The first layer is the embedded layer that uses 5000 length vectors to represent each word and has 40 dimensions.
- Add a dropout layer to perform the dropout layer and reduce overfitting.
- The next layer is the LSTM layer with 200 memory units.
- Output has one neuron, since we have just 2 classes.
- Activation function is sigmoid for binary classification.
- Because it is a binary classification problem, binary_crossentropy is used as the loss function with adam optimizer.

4. Result

They were mixed with various weights after producing the feature vectors.
The results hence have different effects with different models.

4.1 LSTM Model:

The figure given below shows a list of output column predicted by the neural network

```
In [49]: #making predictions for the test dataset

         check=model.predict_classes(X_test)

In [50]: check

Out[50]: array([[0],
                [1],
                [0],
                ...,
                [0],
                [1],
                [0]])
```

We can hence print the fake news from the real news (where 0 depicts fake and 1 depicts real):

```
In [68]: for i in range(0, 10):
             if (check[i][0] == 0):
                 print(test.loc[i])

title    Specter of Trump Loosens Tongues, if Not Purse...
Name: 0, dtype: object
title    #NoDAPL: Native American Leaders Vow to Stay A...
Name: 2, dtype: object
title    Tim Tebow Will Attempt Another Comeback, This ...
Name: 3, dtype: object
title    Pelosi Calls for FBI Investigation to Find Out...
Name: 6, dtype: object
```

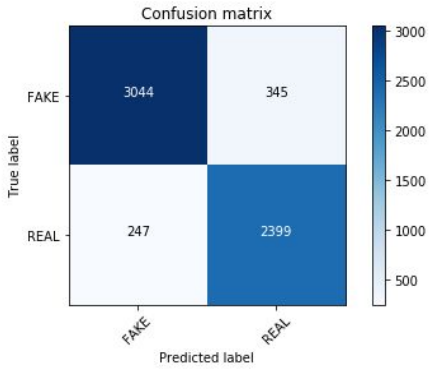
And get a high accuracy score of **90.77%**

```
In [110]: from sklearn.metrics import accuracy_score
          accuracy_score(y_test,y_pred)

Out[110]: 0.907705053852527
```

4.2 Multinomial Model:

```
accuracy:   0.902
Confusion matrix, without normalization
```




```
In [40]: ### Most real
         sorted(zip(classifier.coef_[0], feature_names), reverse=True)[:20]

Out[40]: [(-4.000149156604985, 'trump'),
          (-4.287872694443541, 'hillari'),
          (-4.396389621061519, 'clinton'),
          (-4.899969726208735, 'elect'),
          (-5.176598600897756, 'new'),
          (-5.234730366348767, 'comment'),
          (-5.2739681809973631, 'video'),
          (-5.3868167681180115, 'war'),
          (-5.396821854078974, 'us'),
          (-5.412019714988405, 'hillari clinton'),
          (-5.417137433425386, 'fbi'),
          (-5.48068448454208, 'vote'),
          (-5.566255475855405, 'email'),
          (-5.578238842742501, 'world'),
          (-5.634015380199913, 'obama'),
          (-5.734501455772904, 'donald'),
          (-5.763095255139644, 'donald trump'),
          (-5.785090276725191, 'russia'),
          (-5.846224665218559, 'day'),
          (-5.862110622807369, 'america')]
```

```
In [43]: ### Most fake
         sorted(zip(classifier.coef_[0], feature_names))[:5000]

Out[43]: [(-10.794984555596727, 'abe'),
          (-10.794984555596727, 'abroad'),
          (-10.794984555596727, 'abus new'),
          (-10.794984555596727, 'abus new york'),
          (-10.794984555596727, 'act new'),
          (-10.794984555596727, 'act new york'),
          (-10.794984555596727, 'advic'),
```

Less negative values are most real news, and high negative values hold most fake news.

5. Conclusion

We found that when combined together all 3 aspects are paramount in identifying false news. Using various models including LSTM neural network and Multinomial Naive Bayes Classifier, we obtained the best result with an accuracy of 90.7 percent. Thus, we infer that in deciding whether an article is true or false, linguistic features are key.

6. References

[1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” CoRR, vol. abs/1708.01967, 2017.

[2] S. Feng, R. Banerjee, and Y. Choi, “Syntactic stylometry for deception detection,” in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2, ACL ’12, (Stroudsburg, PA, USA), pp. 171–175, Association for Computational Linguistics, 2012.

[3] N. J. Conroy, V. L. Rubin, and Y. Chen, “Automatic deception detection: Methods for finding fake news,” in Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, ASIST ’15, (Silver Springs, MD, USA), pp. 82:1–82:4, American Society for Information Science, 2015.

[4] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, “Automatic detection of fake news,” CoRR, vol. abs/1708.07104, 2017.

[5] S. Gilda, “Evaluating machine learning algorithms for fake news detection,” in 2017 IEEE 15th Student Conference on Research and Development (SCOReD), pp. 110–115, Dec 2017.

[6] M. Yancheva and F. Rudzicz, “Automatic detection of deception in child produced speech using syntactic complexity features,” in ACL, 2013.

[7] Long, Y. (2017). Fake news detection through multi-perspective speaker profiles. Association for Computational Linguistics.

[8] Wang, William Yang. "" liar, liar pants on fire": A new benchmark dataset for fake news detection." arXiv preprint arXiv:1705.00648 (2017).

[9] Marín, Ignacio Palacio, and David Arroyo. "Fake news detection." Conference on Complex, Intelligent, and Software Intensive Systems. Springer, Cham, 2020.

[10] Roy, Arjun, et al. "A deep ensemble framework for fake news detection and classification." arXiv preprint arXiv:1811.04670 (2018).