Comparing Machine Learning Methods in Fake News Detection

Krystal Dowling
914430
University of Melbourne
krystald@student.unimelb.edu.au

Charlotte Williams
University of Melbourne
914430
email

Coauthor
Affiliation
Address
email

Coauthor
Affiliation
Address
email

Abstract

The abstract paragraph should be indented ½ inch (3 picas) on both the left- and right-hand margins. Use 10 point type, with a vertical spacing (leading) of 11 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract. The abstract must be limited to one paragraph.

1 Introduction

In a world of COVID-19 and an increasingly polarised and unstable political climate, fake news has become a real and credible threat to citizen liberty and health as it continues to disseminate insidious disinformation within a public sphere that increasingly turns to online social-media for news sources [9][10][15]. Within politics specifically it is known to re-enforce confirmation bias of hyper-partisan views, encourage increased political polarisation, and undermines modern-day democracy [7][11].

The task of detecting fake news falls under the umbrella of natural language processing (NLP). NLP is a branch of machine learning which addresses the extraction of semantics and syntactic structure from human language. **this section needs references, will ask laura- also possible include "intuition when solving our problem", not quite sure what this means**. Embeddings were utilised as part of NLP; as they facilitate mapping of text to low-dimensional, learned continuous vector representations that capture meaningful relationships within language for classification [6][12]. Those utilised were chosen based on their appropriateness for the task, and referenced success in related studies, this included TF-IDF [16], GloVe.

Fake news, as a relevant and challenging topic that could benefit from analysis, has a myriad of academic studies on methods of detecting fake news utilising machine learning [1][5][15][16]. In our report, previous studies and results will be taken into consideration to compare statistical machine learning with deep learning techniques. Methods used were those that have been proven to be successful in previous studies; thus for statistical models, a Naïve Bayes classifier was used as a baseline [5] and Support Vector Machine (SVM) as the best performing statistical machine learning model for fake news detection [5] [16]. Furthermore, two variants of deep learning were evaluated for this task: Convolutional Neural Networks (CNN), as they are cited as being appropriate for longer text classification [1], and Recurrent Neural Networks (RNN), as they have been found to be especially effective in sequential text classification [14][15][16]. In order to accurately judge each model and their relevance within the context of society, accuracy, precision, recall, and the generalisability of each model have been considered as evaluation metrics.

2 Data

Due to the lack of a single large and well-rounded dataset of labelled fake and real news, the research conducted as described in this report was performed on a combination of five datasets. Four of the five datasets were sourced from Kaggle, while the final dataset came from Signal Media. After examining the distribution of fake and real news from the Kaggle datasets, a clear imbalance towards fake news was found. To combat this, a sample of true news articles obtained from Signal media were also included, this meant the split between classes in the final dataset was almost even. As inferred from the word cloud shown in Figure X, the news articles revolve primarily around politics, specifically in the United States.

3 Method

3.1 Data Wrangling

From each of the previously listed datasets the article text and class label were extracted to be used in the classification models. Since some of the articles did not have a label, but instead the whole dataset was labelled as 'fake', a label column was created and assigned to these instances in the wrangling process. From this point the datasets were concatenated and treated as a single data frame.

3.2 Preprocessing

LUKE

3.3 Feature Engineering

3.3.1 TF-IDF

Term Frequency Inverse Document Frequency (TF-IDF) vector representations measure the importance of a word based on the number of times it appears in an article against the number of articles the word appears in.

3.3.2 Doc2Vec

LUKE

3.3.3 GloVe

CHARLOTTE

3.3.4 Neural Network Embeddings

Unlike one-hot encoded vectors that are often high-dimensional with uninformed mapping, word embeddings used in neural networks are continuous vectors from learned low-dimensional representations of discrete data that are able to capture relationships within language [8] [6] [12]. While these dense vector representations can be pre-trained, such as with GloVe explained above, they can also be learned within the neural network and thus tailored to training data [3]. The input is required to be integer-encoded for this, which was done using Tokenizer [3]. Then, once input into the embedding layer, the neural network learns optimal weights for these to minimise loss [3] [6].

3.4 Models

3.4.1 Statistical Machine Learning

The baseline model for this research was a Multinomial Naive Bayes. This model was produced using the TF-IDF embeddings, and the top 20,000 word features as determined by the chi-squared test.
Support Vector Machine (SVM) classification was explored as a statistical machine learner for fake news detection. The SVM aims to find a hyperplane which separates instances of different classes with the greatest margin and preferably no incorrect classification. This model was trained separately

Table 1: SVM Results

Embedding	SVM accuracy	F1 Score
TF-IDF	0.8744	0.8495
GloVe	0.7712	0.7230
Doc2Vec	0.7796	0.7476

with three different embedding techniques; TF-IDF, GloVe and Doc2Vec. The TF-IDF vectors underwent feature selection using the chi squared test. After plotting the number of features against the model accuracy, the highest accuracy was achieved using the best 10,000 features. Since the attributes with the GloVe and Doc2Vec embeddings are the dimensions of a single vector, feature selection was not an option as removing any features would change the representation of the vectors. A grid search was then performed to determine the best kernel to fit the data; linear for TF-IDF, polynomial for GloVe and radial basis function for Doc2Vec, were found to produce the best results. The regularisation parameter was kept at its default of 1 to limit any misclassifications. The models were evaluated with the accuracy score and F1 score built into pythons sklearn metrics package.

3.4.2 Convolutional Neural Network

Multiple variations of CNNs were evaluated to get a full idea their effectiveness in fake news detection. A CNN model was employed as it has been found to be effective in NLP problems as it is able to utilise the presence or absence of features, words in this situation, as a distinguishing factor [1]. CNN's comprise of several convolutional layers applied over the input layer and nonlinear activation functions applied to results [8].

The standard model starts with an input layer and an embedding layer (explained above) [3]. Next, the convolutional layer applies different filters to the input, automatically learning the values of its filters based on the task [2]; the kernel value . Next, a max-pooling layer is used to consolidate output while reducing the dimensional complexity and preserving salient information [4] [8] [13]. A flatten layer then converts 3-D data into a vector [4][13] and finally processed by a dense and output layer [3].

This standard model was extended in the study to find the best performing CNN model, with iterations existing that utilised hyperparametrization to optimise values, and the creation multi-channel CNN with two and three channels of the standard CNN model.

3.4.3 Long Short-Term Memory

CHARLOTTE

4 Results

The Multinomial Bayes baseline model performed at an accuracy of 83.10%, with an F1 score of 0.6431.
As shown in Table 1, the SVM achieved its highest accuracy with the TF-IDF embedding. The SVM models with GloVe and Doc2Vec both performed worse than the baseline model, but very similarly to each other. The model with TF-IDF also had a significantly higher F1 score than the other models.

5 Discussion

6 Conclusion

References

[1] F.T. Asr and M. Taboada. "The global rise of "fake news" and the threat to democratic elections in the USA". In: *Public Administration and Policy: An Asia-Pacific Journal* 22.1 (2019), pp. 15–24. DOI: https://doi.org/10.1108/PAP-04-2019-0008.

[2] D. Britz. *Understanding Convolutional Neural Networks for NLP*. 2015. URL: http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/. (accessed: 18.09.2020).

[3] J. Brownlee. *How to Develop a Multichannel CNN Model for Text Classification*. 2018. URL: https://machinelearningmastery.com/develop-n-gram-multichannel-convolutional-neural-network-sentiment-analysis/. (accessed: 13.09.2020).

[4] V. Choubey. *Text classification using CNN*. 2020. URL: https://medium.com/voice-tech-podcast/text-classification-using-cnn-9ade8155dfb9. (accessed: 20.10.2020).

[5] N. Conroy, V. Rubin, and Y. Chen. "Automatic Deception Detection: Methods for Finding Fake News". In: Oct. 2015. DOI: https://doi.org/10.1002/pra2.2015.145052010082.

[6] W. Koehersen. "Neural Network Embeddings Explained". In: *towards data science* (2018).

[7] T. Lee. "The global rise of "fake news" and the threat to democratic elections in the USA". In: *Public Administration and Policy: An Asia-Pacific Journal* 22.1 (2019), pp. 15–24. DOI: https://doi.org/10.1108/PAP-04-2019-0008.

[8] M.M. Lopez and J. Kalita. *Deep Learning applied to NLP*. 2017. eprint: 1703.03091.

[9] G. Nyilasy. "Fake News in the Age of COVID-19". In: *Inside Business* (2020).

[10] A. Rajan. "Fake news: Too important to ignore". In: *BBC* (2017).

[11] M. Rosenwald. "Making media literacy great again". In: *Columbia Journalism Review* (2017).

[12] R. Ruizendaal. "Deep Learning 4: Why You Need to Start Using Embedding Layers". In: *towards data science* (2017).

[13] S. Saha. "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way". In: *towards data science* (2018).

[14] E. Saravia. *Deep Learning for NLP: An Overview of Recent Trends*. 2018. URL: https://medium.com/dair-ai/deep-learning-for-nlp-an-overview-of-recent-trends-d0d8f40a776d. (accessed: 15.09.2020).

[15] K. Shu et al. "Fake News Detection on Social Media: A Data Mining Perspective". In: *SIGKDD Explor. Newsl.* 19.1 (2017), pp. 22–36. DOI: https://doi.org/10.1145/3137597.3137600.

[16] N. Smitha and R. Bharath. "Performance Comparison of Machine Learning Classifiers for Fake News Detection". In: *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. 2020, pp. 696–700. DOI: 10.1109/ICIRCA48905.2020.9183072.