Identifying Characteristic Features of Fake News Articles for Deep Learning-Based Identification

Arjun Sharma

Abstract: Fake news articles rapidly spread online, spreading misinformation, and weakening democracy and credible journalism. This research identifies characteristic stylistic features in the text of fake news articles to build a deep learning-based classifier of news articles, including topics, sentiment, and length of the titles and bodies of articles.

An experimental approach is taken to compare the effectiveness of multiple feature selections as input data for the binary classification of fake news articles by a neural network.

The top-performing feature selection and neural network architecture result in a classifier that achieves 89.7% accuracy on a testing set.

1. Introduction

The challenge posed by the spread of fake news that is published and exchanged over the internet is rapidly becoming increasingly consequential. False information spread over social media has been found to have significant implications on real-life human behavior, resulting in consequences on public health, as seen in regard to unfounded skepticism about the MMR vaccine. It is estimated that fake news and information online was responsible for half of the 3% fall in MMR vaccine coverage in certain areas between 2014 and 2018 [1]. Additionally, it may also have an impact on the political climate of countries [2], as seen in an instance where the false 'Pizzagate' story eventually resulted [3] in a potentially violent encounter in a public setting.

However, there remains a general lack of ability to distinguish between reliable and unreliable sources of information. A survey by Stanford University in 2016 found the ability of a group of middle school, high school, and college students to identify false information to be so poor that democracy could be 'threatened' because of the ease at which this allows disinformation to 'spread and flourish' [4]. A 2016 Pew Research study found that even though 39% of American adults are very confident and another 45% are somewhat confident in their ability to detect fake news, 23% of adults have still knowingly or unknowingly contributed to the spread of misinformation by sharing fake news online [5].

Addressing this challenge by manually identifying fake news, however, is impractical. Labeling fake news on content-sharing platforms poses two challenges: firstly, identifying which kinds of content must be targeted for Labeling, and secondly, the near impossibility of appropriately tagging such content without being affected by personal biases to ensure user trust in this moderation is not eroded [6].

The purpose of this paper is to identify the set of stylistic natural language features which serve as the most effective input features for a deep learning-based binary classifier of textual fake news items.

2. Related works

Zhou et al. categorized techniques to perform automated binary classification of news articles under five categories [7]:

1. Knowledge-based approaches, which use automated fact-checking processes to assess the authenticity of news by comparing unverified news articles to verified facts.
2. Style-based approaches, in which it is presumed that malicious entities that publish fake news prefer to write it in a particular style to achieve their intended effect, therefore creating a set of quantifiable characteristics that can well represent fake news content.
3. Propagation-based approaches, in which the propagation of fake news over social media is represented in a tree-like structure
4. Source-based approaches, in which the credibility of the source publishing the news story is assessed
5. Hybrid approaches, which combine different aspects of the above approaches.

This paper employs a style-based method to perform binary classification of fake news articles.

3. Method

A dataset consisting of 20 800 articles was obtained from Kaggle [8]. In it, 10387 articles were flagged as unreliable (labeled as 0), and the remaining 10413 were flagged as reliable (labeled as 1). The title, author name, and body text were also present for each article.

After cleaning the dataset through lemmatization, lowercasing, and stop-word removal, three categories of stylistic features are extracted for article titles and bodies.

a. **Topics**: Two separate Latent Dirichlet Allocation (LDA) models are used to identify topics in both the titles and bodies. Each model identified 20 topics in its respective corpus.

All tokens in both corpora are given a probability of being associated with each of the 20 topics of that corpus. The probability of a document (a title or body) being associated with a topic is found by calculating the average association of its tokens with that topic. A probability vector is stored for each document, where each of its 20 values represents the probability of the document being associated with the corresponding topic [9].

To analyze the relationship between topics and article reliability, the topics most frequently covered by an author can be compared to the average reliability of that author. Their reliability is calculated as the fraction of articles written by the author that are labeled as reliable. The probability of writing about a topic is calculated as the average probability of their documents being associated with it.

Table 1 compares author reliability to the average probability they discuss a particular topic in their article body.

Scatterplot	Tokens	Trend
	Game	Tend to write reliable articles
	Team	
	Player	
	Season	
	Last	

Table 1: Relationship between author topics and reliability

b. **Sentiment**: The VADER sentiment analyzer [10] is used to extract sentiment scores for the titles and bodies. It returns a compound score representing the overall sentiment of a document based on its preconfigured lexicon, ranging from -1 (negative) to 1 (positive); as well as ratios representing the proportions of words in the text which fall under positive, negative, and neutral thresholds in its lexicon.

Table 2 compares sentiment polarity scores of reliable and unreliable articles for article titles and bodies.

Sentiment score	Title		Body	
	Reliable	Unreliable	Reliable	Unreliable
Compound	-0.139	-0.165	0.175	0.00309
Positive (%)	9.23	11.5	8.65	8.90
Negative (%)	15.9	20.2	7.62	8.37
Neutral (%)	74.9	68.3	83.7	82.7

Table 2: Comparison of mean sentiment scores

Two trends emerge here:

1. Unreliable articles tend to use less neutral language in their titles, compared to reliable articles, by 6.6% on average
2. Reliable articles tend to be significantly more positive on average, with an average compound sentiment score greater than unreliable articles by approximately 0.144.

   c. Length: Titles of reliable articles tend to be longer than those of unreliable articles, with their median length being 16 characters longer. Reliable article bodies are significantly longer than unreliable article bodies by ~95%. Table 3 compares statistical features of the length of reliable and unreliable articles for article titles and bodies.

Length	Title		Body	
	Reliable	Unreliable	Reliable	Unreliable
Mean	79.6	65.2	5214.1	3875.9
Median	81	65	4591	2351
Standard deviation	15.6	31.0	4313.8	5753.6

Table 3: Relationship between author topics and reliability

4. Experimental Design

The previous section gives six groups of features for each article:
1. Title topic probability vectors, consisting of 20 individual probability values
2. Body topics probability vectors, consisting of 20 individual probability values
3. Title sentiment polarity scores, consisting of 4 individual polarity scores
4. Body sentiment polarity scores, consisting of 4 individual polarity scores
5. Title length, a single whole number
6. Body length, a single whole number

The number of articles for which topics, sentiment, and length are available is 9698. This set is split into training and testing sets, consisting of 90% and 10% of the articles. Models are trained with a 15% validation split.

Multiple neural networks are trained to identify the combination features that most effectively characterize fake news articles when used as input features for a binary classification model, each of which uses one of the following feature selections:

1. All features
2. Title features (its topics, sentiment, and length)
3. Body features (its topics, sentiment, and length)
4. Six different combinations, each of which uses all but one of the six groups of features

To find the optimal hyperparameters for each of the nine neural networks, the Keras tuner [11] with a hyperband search algorithm trains different variations of a base neural network architecture:

1. A dense input layer consisting of 12 neurons
2. 2-20 hidden layers, each of which consists of 32-512 neurons which use a relu activation function. The exact number of hidden layers and neurons is decided by the hyperparameter tuner for each iteration of the model.
3. An output layer that uses a sigmoid activation function to predict a probability between 0 (reliable) and 1 (unreliable) [12] for the article

Each variation is trained for ten epochs. After the search is complete, a model with the same hyperparameters as the best-performing combination is trained for 50 epochs and saved

5. Measurements

The nine models can be compared by evaluating their performance when used to predict articles in the testing set in terms of in two metrics.

1. Their accuracy, i.e., the percentage of correctly predicted examples in the testing set
2. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve of the model, which compares the true positive rate (y-axis) to the false positive rate (x-axis) of the classifier at all possible classification thresholds [13]

Table 4 lists model accuracy and AUC for each of the nine models when used to make predictions on articles in the test set.

Model No.	Features	Test Accuracy (%)	AUC
1	All features	88.7	0.957
2	Title features	80.0	0.865
3	Body features	85.2	0.934
4	Excluding title topics	84.7	0.938
5	Excluding title sentiment	88.9	0.956
6	Excluding title length	89.9	0.961
7	Excluding body topics	80.0	0.869
8	Excluding body sentiment	87.6	0.959
9	Excluding body length	88.7	0.958

Table 4: Comparison of performance of optimized neural networks for each model feature combination

6. Discussion

First, we compare the importance of title features and body features. Fig 1. contains ROC curves for each of the first three models.
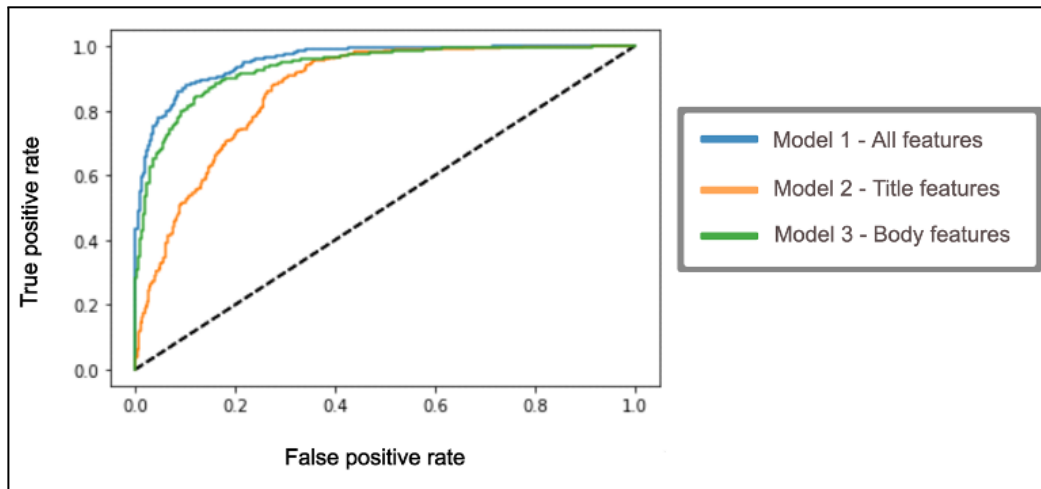

Figure 1: Comparison of ROC curves for models 1, 2, and 3

The ROC plot in Fig. 1 indicates that body features create a more accurate classifier than title features. However, the best performing of the first three models is model 1, which achieves 88.7% accuracy on the testing set.

We then compare the importance of each of the six features by comparing the performance of model 1 with models which include all but one selected feature. Fig. 2 compares the ROC curve of the first model with the ROC curves of models 4-9.
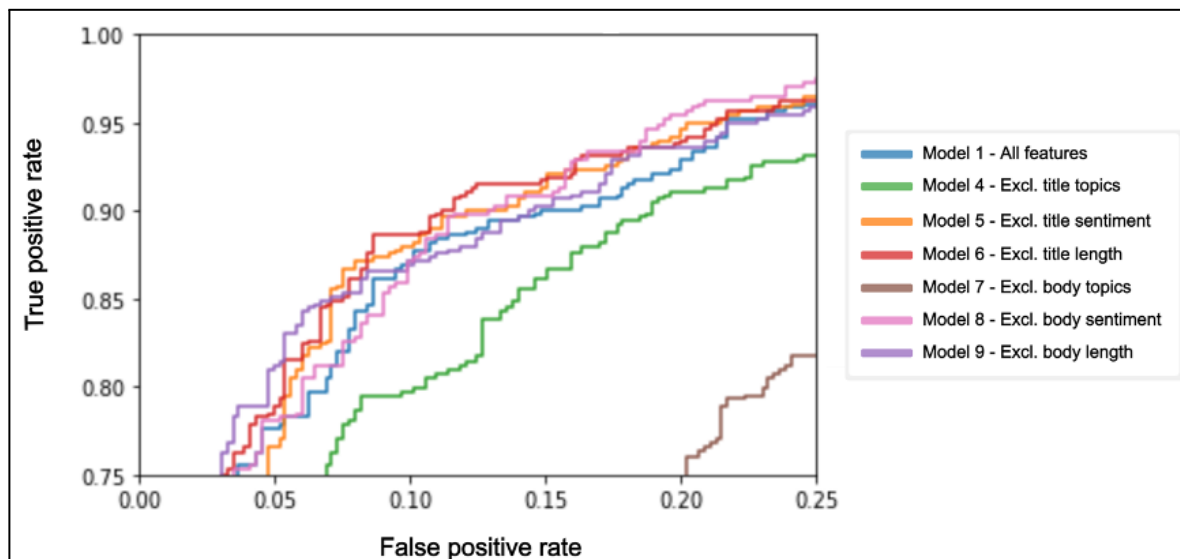
Figure 2: Comparison of ROC curves of model 1 and models 4-9, zoomed in to the top left of the graph

Fig 2. indicates that performance drastically falls when title and body topics are excluded. It also suggests a drop in performance when body sentiment is excluded.

However, performance appears to improve when title sentiment is excluded (accuracy rises by 1.1% relative to model 1), as well as when title length is excluded (accuracy increases by 0.2%). Body length appears to have no statistically significant impact on the performance of the model.

5. Conclusion

The above findings indicate that the body of an article is a better predictor of the reliability of an article than its title.
Furthermore, some stylistic features of the titles and bodies of articles are characteristic of fake news articles: title topics, body topics, and body sentiment. However, title sentiment and the length of both the title and body are not valid predictors of the reliability of an article.

The best-performing neural network achieves 89.9% accuracy in predicting the reliability of articles in the test set. It uses the title topic vectors, body topic vectors, and body sentiment polarity scores as its input features.

Future research using style-based techniques may use morphological analysis, such as part-of-speech tagging. Emerging research with transformer-based natural language classifiers may yield higher accuracy on more complex datasets.

6. References

1. Meschi, Meloria, David Eastwood, and Ravi Kanabar. "The Real-World Effects of 'Fake News' – and How to Quantify Them." SCL, August 10, 2020. www.scl.org/articles/12022-the-real-world-effects-of-fake-news-and-how-to-quantify-them.
2. Lee, Terry. "The Global Rise of 'Fake News' and the Threat to Democratic Elections in the USA." emeraldinsight.com, March 18, 2019. www.emerald.com/insight/content/doi/10.1108/PAP-04-2019-0008/full/pdf.
3. WPSU - Penn State Public Media. "Case Study – Fake News Dissemination: Pizzagate (Continued)." Case study – fake news dissemination: Pizzagate (continued). The Arthur W. Page Center. Accessed October 22, 2021. www.pagecentertraining.psu.edu/public-relations-ethics/introduction-to-the-ethical-implications-of-fake-news-for-pr-professionals/lesson-2-fake-news-content/case-study-fake-news-dissemination-pizzagate- continued/.
4. "Evaluating Online Information: The Cornerstone of Civic Online Reasoning." Stanford History Education Group, November 22, 2016. stacks.stanford.edu/file/druid:fv751yt5934/SHEG%20Evaluating%20Information%20Online.pdf.
5. Barthel, Michael, Amy Mitchell, and Jesse Holcomb. "Many Americans Believe Fake News Is Sowing Confusion." Pew Research Center's Journalism Project. Pew Research Center, December 15, 2016. www.pewresearch.org/journalism/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/.

6. Stewart, Elizabeth. "Detecting Fake News: Two Problems for Content Moderation." Philosophy & Technology. Springer Netherlands, February 11, 2021. link.springer.com/article/10.1007/s13347-021-00442-x.

7. Zhou, Xinyi, and Reza Zafarani. "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities." ACM Comput. Surv. 1, July 17, 2020. arxiv.org/pdf/1812.00315.pdf.

8. UTK Machine Learning Club. "Fake News". Retrieved from kaggle.com/c/fake-news/data

9. Řehůřek, R., & Sojka, P. (2010, Μάιος 22). Software Framework for Topic Modelling with Large Corpora. Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, 45–50. Valletta, Malta: ELRA

10. Hutto, C.J. & Gilbert, E.E. (2014). "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text". Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014

11. Chollet, F., & Others. (2015). Keras. keras.io

12. Google. (n.d.). Machine learning glossary. Google Developers. developers.google.com/machine-learning/glossary

13. Google (n.d.) Classification: ROC Curve and AUC. Google Developers. developers.google.com/machine-learning/crash-course/classification/roc-and-AUC