Fake News detection using RL and Neural Networks

Ezana N. Beyenne  
Yunze Chen  
Tanay Jain  
Yuan Zhou

MSDS 464: Final Project

## Professor Shreenidhi Bharadwaj

.

03/07/2021

**Abstract**

Fake news (also known as fabricated news) is typically found in traditional news, social media, or fake news websites, presenting news as factually accurate even though it has no basis.[[1]](#footnote-1) False, distorted, or exaggerated information result in bad decisions or wrong opinions that neglect to have a legitimate and factual basis. Experiments show that using machine learning and the development of a neural network could distinguish legitimate news from fake news. However, the use of artificial intelligence challenges how information is filtered and organized. Research has shown machine learning algorithms could easily interpret positive statements but could not interpret negative statements. To correctly identify fake news for machine learning algorithms, we must somehow weed out human bias and prejudices during its training phase.[[2]](#footnote-2) The neural network algorithms have shown that they are effective in combating fake news. Fake news spreads like wildfire, so we are seeing that using Reinforcement learning can be used to help labeling fake news to assist train the neural network models as quickly as possible.[[3]](#footnote-3)

**Introduction and Problem Statement**

Fake news (also known as fabricated news) is typically found in traditional news, social media, or fake news websites, presenting news as factually accurate even though it has no basis in fact.1 False, distorted, or exaggerated information neglect to have any legitimate and factual basis, and therefore people make decisions and form wrong opinions. Machine learning has been used in the past to try and distinguish legitimate news from fake news. However, research has shown machine learning algorithms could easily interpret positive statements but could not interpret negative statements. To correctly identify fake news for machine learning algorithms, we must somehow weed out human bias and prejudices during its training phase.[[4]](#footnote-4) The development of a neural network could be the answer, but artificial intelligence poses challenges in how information is filtered and organized.

Information shapes our view of the world and allows us to make crucial decisions based on our information. Information we receive allows us to form ideas about people or situations around us. Fake news has led to bullying, violence against innocent people, racist ideas, and fear-mongering. Fake news significantly impacted the last American presidential election.2 Fabricated news has enormous popular appeal, are unsubstantiated stories, yet consumed by millions of people. Unfortunately, these fabrications are not limited to politics and exist within news concerning vaccination, nutrition, and stock values. Claire Wardle identifies seven types of fake news.[[5]](#footnote-5)   
1. *Satire or parody*: material with the potential to fool, but no intention to harm.  
2. *False connection*: material with visuals and headlines that do not match the content.  
3. *Misleading content*: material intended to frame an issue or individual.

4. *False context*: material whose real content is shared with false contextual information.

5. *Impostor content*: material whose genuine sources are impersonated with false, made-up stories.

6. *Manipulated content*: the material with genuine information or images are doctored up to deceive.

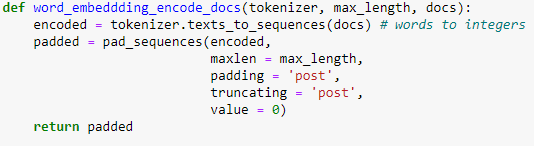
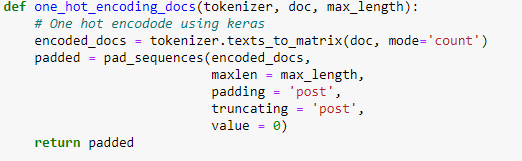
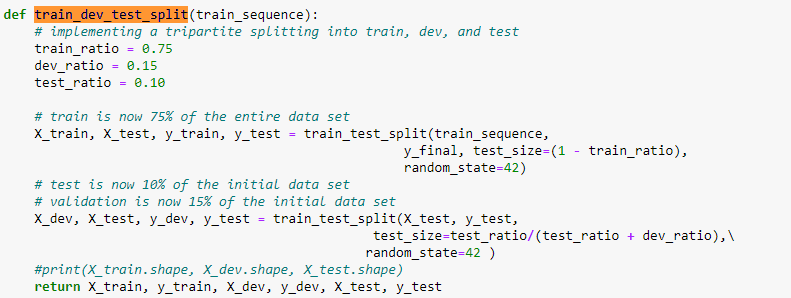
7. *Fabricated content*: the material is intended to deceive or harm with 100 percent fake content.

**Review of Literature**

Research using machine learning was a tool anticipated to identify fake news and prevent them from going viral and spreading misinformation.3 However, machine learning algorithms do a poor job of identifying fake news. They can even generate fake news without any human intervention.Research conducted by two MIT doctoral students found that computers could identify machine learning generated text but could not identify if that text was true or false. Machine learning algorithms could easily interpret positive statements but could not interpret negative statements.2 The database used to train machine learning algorithms, called Fact Extraction and Verification (FEVER), had some inherent biases when trying to identify fake news. Kaggle, world’s largest data science community, conducted a Fake News challenge where competitors attempted to develop a machine learning program to identify articles that might be unreliable fake news articles.

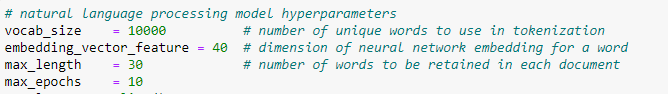
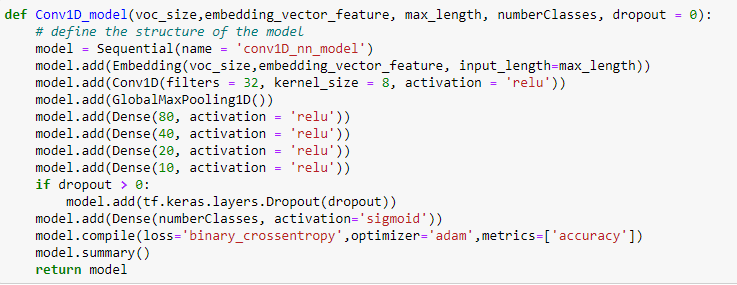
Fake news detection using various Natural Language Processing (NLP) and machine learning is still an active research area with most of the focus being on social media platforms. People have moved from traditional print and stand-alone websites to social media, such as Twitter, Facebook and others, to consume news.[[6]](#footnote-6)

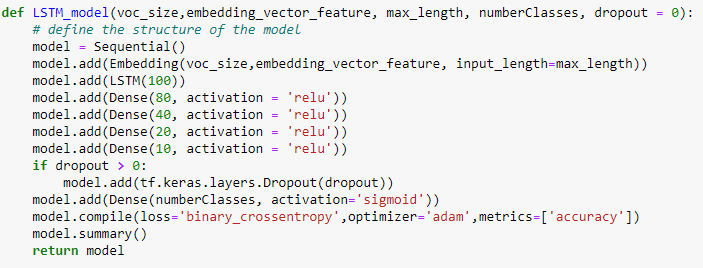
**Neural Network Methods**

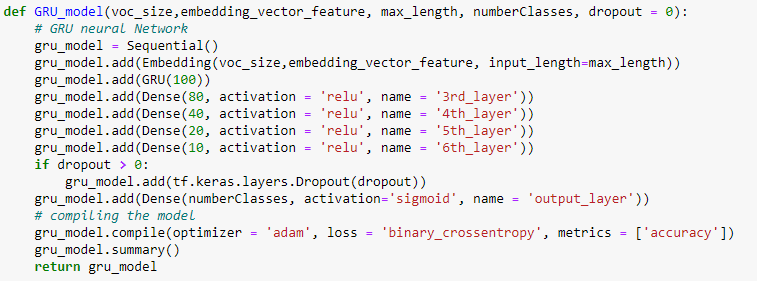
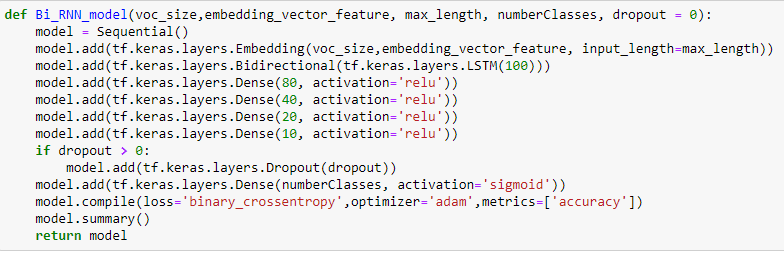
Various neural network algorithms can correctly identify whether an article was fake news or not in the Fake News Kaggle challenge's training dataset. I converted the data in train.csv into two vectorization techniques, namely word embedding and one-hot encoding, to compare their effectiveness on the 1D CNN, LSTM, GRU, and Bidirectional LSTM neural network methods. I then employed a tripartite splitting of the train.csv data to get a 75/15/10 percent train/dev/test split on nearly 18,250 rows of data after cleanup.   
 **Table 1.** Word Embedding code  
  
**Table 2.** One hot embedding code  
  
**Table 3**. Train dev test split code

I used a factorial design with the alternative methods mentioned above, using both word embedding and one-hot encoding vectorization techniques. In addition, I included alternative settings to the neural network methods by adding dropout regularizations. Both vectorization techniques include the dropout regularization of 25% as well as 50 % on the neural networks 1D CNN, LSTM, GRU, and Bidirectional LSTM.

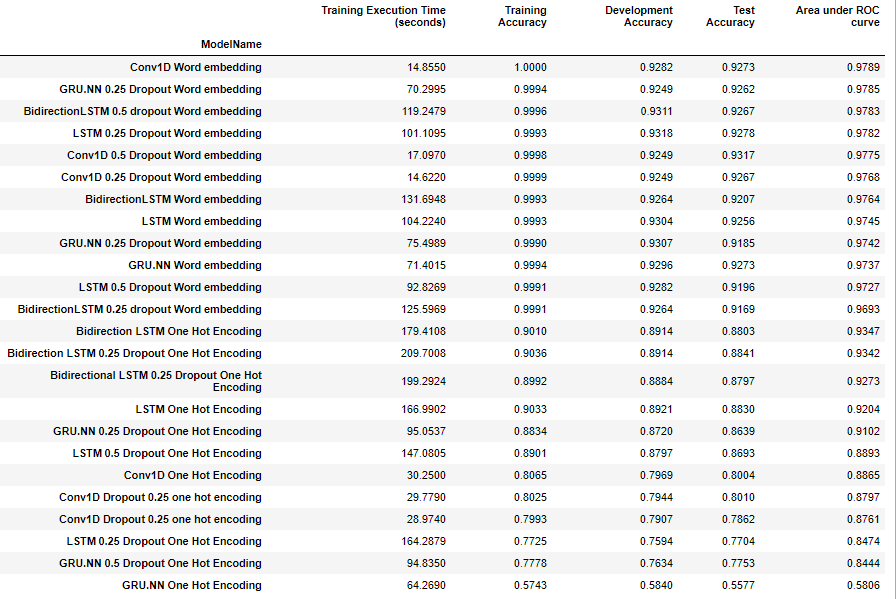
Early stopping is employed to monitor the neural networks' performance during the training phase to reduce over-fitting and improve generalization. I provided graphs of training and development that set accuracy and loss and graphs of the area under the ROC curve. Since this was a binary classification problem ( fake news or legitimate news), I employed the sigmoid activation function in the neural networks' output layer, with a loss of binary cross-entropy.

The initial area under the curve score (AUC) was 0.5. To improve the score, I adjusted the hyperparameters shown in Table 4 and increased the number of hidden layers in the neural networks. After several attempts to find the ideal number of epochs with the early stopping taking place, I settled on ten epochs. The GRU neural network with one-hot encoding could have used more hidden layers since it still had an AUC score around 0.5, but I hesitated because the GRU with word embedding had a higher score.   
**Table 4.** Hyperparameters used in the neural networks  
 I used Keras Tensorflow 2.1.0 to build the four neural network functions while including dropout as an optional parameter. The tables below show various neural networks' setup because of numerous trials to find the right combinations of layers and structures.  
**Table 5.** Convolution 1D model

  
**Table 6:** LSTM model

  
**Table 7:** GRU model  
  
**Table 8:** Bidirectional LSTM

**Results**

The word vectorization technique has the best performance with the word embedding compared to the one-hot encoding. Additionally, results in Table 9 show that one-hot encoding has the longest training execution time but lacks top test accuracy and area under the curve results. Surprisingly, the Conv1D word embedding model has the top area under the ROC curve and test accuracy scores.  
**Table 9.** Results of all the models

**Weak Supervised RL Methods**

<https://arxiv.org/abs/1912.12520>

**RL Code**

<https://github.com/yaqingwang/WeFEND-AAAI20>

**Results**

**Conclusion**

**Works Cited**

Grothaus, Michael. “*Machine Learning Isn't Effective at Identifying Fake News*.” Fast Company, October 15, 2019. <https://www.fastcompany.com/90417625/machine-learning-isnt-effective-at-identifying-fake-news>.

Hamdi, Tarek*. Top Research About Fake News Detection 2019*. Retrieved from  
<https://www.kaggle.com/c/nlp-getting-started/discussion/123454>

Kaggle. *Fake-news*. Retrieved from <https://www.kaggle.com/c/fake-news>

Quebec, Journalists du. “*Impacts of Fake News.*” 30sec to check it out. Fédération professionnelle des journalistes du Québec, January 10, 2019. <https://30secondes.org/en/module/impacts-of-fake-news/>.

Parikh, Shivam B, Vikram Patil, and Pradeep K Atrey. *On the Origin, Proliferation and Tone of Fake News*. IEEE Conference Publication, April 25, 2019. <https://ieeexplore.ieee.org/document/8695387>.

Tufekci, Zeynep. “*It's the (Democracy-Poisoning) Golden Age of Free Speech.*” Wired. Conde Nast, January 16, 2018. <https://www.wired.com/story/free-speech-issue-tech-turmoil-new-censorship/>.

Vega, Nicolas. “*Facebook 'Peddling in an Addictive Drug Called Anger': Steve Jobs Adviser.*” New York Post. New York Post, June 15, 2020. <https://nypost.com/2020/06/12/facebook-peddling-in-an-addictive-drug-called-anger-steve-jobs-advisor/>.

Wardle, Claire. “*Fake News. It's Complicated.*” First Draft, May 15, 2017. <https://firstdraftnews.org/latest/fake-news-complicated> .

Wang, Yaqing, Weifeng Yang, Fenglong Ma, Jin Xu, Bin Zhong, Qiang Deng, and Jing Gao. “Weak Supervision for Fake News Detection via Reinforcement Learning.” arXiv.org, January 20, 2020. <https://arxiv.org/abs/1912.12520v2>

Tschiatschek, Sebastian, Adish Singla, Manuel Gomez Rodriguez, Arpit Merchant, and Andreas Krause. “Fake News Detection in Social Networks via Crowd Signals.” Arxiv.org, March 2, 2018. <https://arxiv.org/pdf/1711.09025.pdf>

1. Zeynep Tufekci, “It's the (Democracy-Poisoning) Golden Age of Free Speech,” Wired (Conde Nast, January 16, 2018), <https://www.wired.com/story/free-speech-issue-tech-turmoil-new-censorship/>. [↑](#footnote-ref-1)
2. Journalists du Quebec, “Impacts of Fake News,” 30sec to check it out (Fédération professionnelle des journalistes du Québec, January 10, 2019), <https://30secondes.org/en/module/impacts-of-fake-news/>. [↑](#footnote-ref-2)
3. Wang, Yaqing, Weifeng Yang, Fenglong Ma, Jin Xu, Bin Zhong, Qiang Deng, and Jing Gao. “Weak Supervision for Fake News Detection via Reinforcement Learning.” arXiv.org, January 20, 2020. <https://arxiv.org/abs/1912.12520v2>

   [↑](#footnote-ref-3)
4. Michael Grothaus, “Machine Learning Isn't Effective at Identifying Fake News,” Fast Company (Fast Company, October 15, 2019), <https://www.fastcompany.com/90417625/machine-learning-isnt-effective-at-identifying-fake-news>. [↑](#footnote-ref-4)
5. Claire Wardle, “Fake News. It's Complicated.” First Draft, May 15, 2017, <https://firstdraftnews.org/latest/fake-news-complicated/>. [↑](#footnote-ref-5)
6. Shivam B Parikh, Vikram Patil, and Pradeep K Atrey, “On the Origin, Proliferation and Tone of Fake News,” On the Origin, Proliferation and Tone of Fake News (IEEE Conference Publication, April 25, 2019), <https://ieeexplore.ieee.org/document/8695387>. [↑](#footnote-ref-6)