Fake News Detection

Vatsalya Anand

Abstract

This comprehensive study aims to investigate the effective detection of fake news by utilizing the widely acclaimed Liar dataset in conjunction with three robust models: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Bidirectional Long Short-Term Memory (BiLSTM). The results of this study reveal that while the CNN model exhibits exceptional accuracy, the BiLSTM model showcases superior precision and recall. These findings underscore the potential of these advanced models in effectively identifying and combatting the pervasive issue of misinformation in the ever-evolving digital landscape. Thus, this study makes a valuable contribution towards the ongoing efforts to mitigate the spread of fake news in the digital age.

Keywords: fake news, detection, Liar dataset, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), accuracy, precision, recall, misinformation, digital landscape.

I. INTRODUCTION

Researchers and practitioners are actively engaged in extensive research and development efforts to tackle the rampant spread of misinformation through the creation of automated systems for fake news detection. The primary objective of these endeavors is to identify and categorize news articles or statements as either genuine or fabricated. To accomplish this task, various cutting-edge machine learning techniques, including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM), are commonly employed. These techniques have proven to be highly effective in this domain. Additionally, the widely recognized Liar dataset serves as a valuable resource for training and evaluating the performance of fake news detection models. While these models exhibit promising capabilities in identifying and combatting fake news, it is crucial to acknowledge that each approach possesses its own set of strengths and limitations. In light of this, the present study aims to make a meaningful contribution to the fight against fake news by extensively exploring the effectiveness of CNN, LSTM, and BiLSTM models, leveraging the Liar dataset as a foundational benchmark.

II. DATASET

The LIAR dataset is open-access and designed to detect false information. Around 12.8 thousand brief statements were manually verified. POLITIFACT.COM, a fact-checking website, provided the dataset with detailed analysis reports and source document links.

The LIAR dataset includes veracity, subject, context, speaker, position, political affiliation, and historical timeframe for each statement. The rich metadata lets researchers analyze statements and identify misleading information.

The LIAR dataset's data engineering preprocessing steps and techniques are not disclosed. Data cleansing, standardization, and feature extraction are usually done in such datasets to ensure uniformity and speed up model training and evaluation.

The LIAR dataset is widely used in fake news detection research to evaluate models and methods. Researchers created and tested CNN, LSTM, and BiLSTM machine learning models to identify fake news using the dataset.

To conclude, the LIAR dataset is invaluable for false information research. The platform provides a large collection of labeled statements from POLITIFACT.COM to help researchers study misleading information and build effective models for detecting it. While the dataset's data engineering details are not stated, standard preprocessing techniques were likely used to ensure data quality and facilitate model training and evaluation.

Dataset Statistics	
Training set size	10,269
Validation set size	1,284
Testing set size	1,283
Avg. statement length (tokens)	17.9
Top-3 Speaker Affiliations	
Democrats	4,150
Republicans	5,687
None (e.g., FB posts)	2,185

III. MODEL DESCRIPTION

Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM) are commonly used models for identifying fake news. These models employ deep learning techniques to comprehend the meaning, order, and context of text data.

The CNN model utilizes layers to extract features and reduce dimensions. The LSTM model makes use of recurrent neural networks to capture the sequence of statements. The BiLSTM model enhances the LSTM model by considering both past and future contexts.

Researchers have evaluated the performance of these models on the LIAR dataset to assess their effectiveness in detecting fake news. The CNN model demonstrates the highest accuracy, while the BiLSTM model exhibits better precision and recall. LSTM models also exhibit promise in the detection of fake news.

The selection of a model depends on factors such as the dataset, research question, and performance metrics. Researchers can experiment with different models and techniques to determine the most suitable one for their specific use case.

In summary, CNN, LSTM, and BiLSTM models are widely employed for detecting fake news using the LIAR dataset. These models leverage deep learning to comprehend the meaning, order, and context of text data. Researchers have the flexibility to explore various approaches to identify the most effective one for their use case.

IV. DATA LOADING AND PREPROCESSING

First, load the Liar dataset into your system to initiate the process. This dataset comprises a collection of sentences that will be utilized for subsequent analysis.

Then, it is crucial to apply tokenization to each sentence in the dataset. Tokenization is a vital step in natural language processing that involves breaking down a sentence into its individual components, such as words or phrases.

The tokenization process consists of the following steps:

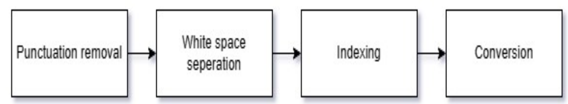
1. Punctuation removal: Eliminate all punctuation marks from the sentence to focus solely on the words or phrases.

2. White space separation: After removing punctuation, convert the sentence into a list of words. Each word is separated by white spaces, facilitating analysis and processing.

3. Indexing: Assign a unique index to each distinct word in the sentence. This indexing aids in identifying and referencing specific words in the dataset.

4. Conversion: Ultimately, replace each word in the sentence with its corresponding unique index. This step is crucial for subsequent analysis and modeling of the dataset.

By following these steps, we can confidently tokenize the sentences in the Liar dataset and prepare them for further analysis and processing.



In the data extraction process, the dataset is divided into independent variables (statement column) and target variable (Label column). The label column is then encoded into binary values (0 or 1) to determine if the news is fake or real.

IV. LOSS FUNCTION

The selection of the loss function in models designed to identify fake news is of utmost importance in determining the efficacy of training and the overall performance of the models. Several widely used loss functions in this context include binary cross-entropy, categorical cross-entropy, and mean squared error.

Binary cross-entropy is commonly employed in binary classification tasks, where the objective is to categorize statements as either true or false. Conversely, categorical cross-entropy is utilized in multi-class classification tasks, where statements are categorized into distinct groups such as true, mostly true, half true, barely true, false, or even pants on fire.

Researchers are investigating novel approaches in addition to traditional loss functions. An effective strategy involves implementing focal loss, where greater emphasis is placed on challenging or incorrectly classified samples. This directs the model's attention towards these difficult instances. An alternative strategy involves integrating adversarial loss, which seeks to improve the model's resilience against potential attacks.

The choice of a suitable loss function is contingent upon several factors, such as the problem at hand, the data that is accessible, and the intended behavior of the model. Consequently, researchers frequently carry out experiments to determine the most appropriate loss function that can produce optimal performance.

Ultimately, the selection of a loss function in models designed to identify fake news is heavily contingent upon the specific goals and attributes of the given problem. Although binary cross-entropy and categorical cross-entropy are widely utilized, researchers persist in investigating and creating novel loss functions to enhance the effectiveness of these models and address the difficulties related to fake news detection.

V. OPTIMIZATION ALGORITHM

The Adam optimizer is widely used for optimizing deep learning models, including those used for detecting fake news. It is an extension of the stochastic gradient descent (SGD) algorithm that incorporates adaptive learning rates and momentum. Adam, short for Adaptive Moment Estimation, combines the benefits of two other optimization algorithms, AdaGrad and RMSProp. This algorithm maintains a dynamic learning rate for each parameter, adjusting it based on gradient magnitude. It also incorporates momentum by tracking the exponentially decaying average of previous gradients. The use of adaptive learning rate and momentum in Adam enables faster convergence and helps prevent getting stuck in local minima. It is particularly suitable for models with large datasets and high-dimensional parameter spaces.

While Adam is widely used, researchers have the opportunity to explore alternative optimization algorithms to find the best performance for their specific use case. Other optimization algorithms include stochastic gradient descent (SGD), AdaGrad, RMSProp, and Nesterov accelerated gradient (NAG). Researchers can investigate modifications or alternatives to improve convergence speed, stability, and overall generalization performance. This may involve implementing adaptive learning rate schedules, weight decay techniques, or gradient clipping mechanisms. The choice of optimization algorithm depends on factors such as the model structure, dataset size, and available computational resources. Researchers often experiment with different optimization algorithms to determine the most effective one for detecting fake news.

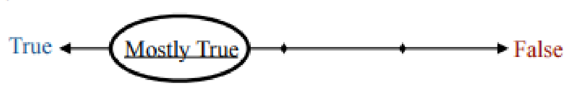
VI. METRICS AND EXPERIMENTAL RESULTS

Two approaches were taken to get the results.

1. 6-way classification
2. 2-way classification

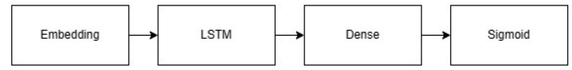
For a 6-way classification the labels were taken as it is. Ranging from Pants on fire to False to True.

While in 2-way classification we made the label into either True or False. Like 'barely true' was taken as False; 'half-true' as True etc.

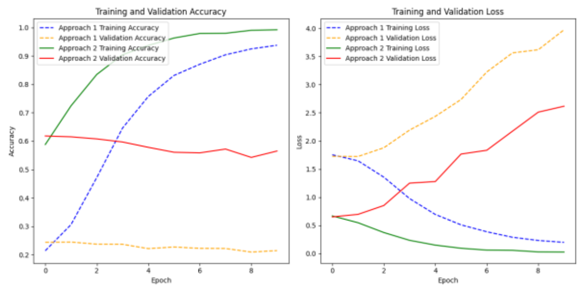


LSTM

LSTM is a type of Recurrent Neural Network, It effectively captures sequential dependencies in text, models long-term relationships, adapts to variable-length sequences, and learns features indicative of misinformation. Its memory cell aids in retaining important context over longer sequences, making it suitable for understanding the historical context within news articles.



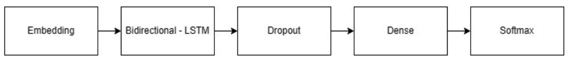
<u>Results LSTM:</u>



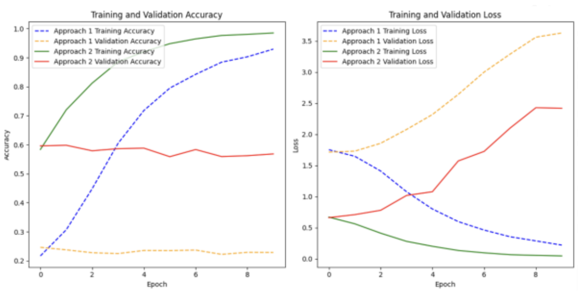
As discussed earlier we have used two approaches to evaluate accuracy and loss, as shown in the figures above. Approach 1 is 6-way classification while the approach 2 is 2-way classification. It is evident from the graph that approach 2 has the highest accuracy, while approach 1 has a lower validation accuracy. Also approach 1 is highly overfit while approach 2 we were able to minimize the overfitting to the maximum extent. We can also see from the loss curve that approach 2 has lower loss as compared to approach 1 which indicates better performance.

BiLSTM

BiLSTM stands for Bidirectional Long Short-Term Memory. It is a type of recurrent neural network (RNN) that processes input sequences in both forward and backward directions, allowing the network to capture context from both past and future inputs. This makes BiLSTM particularly useful for tasks such as natural language processing and speech recognition, where the meaning of a word can depend on the words that come before and after it.



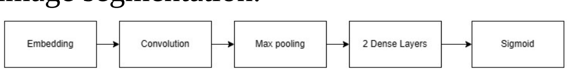
Results BiLSTM



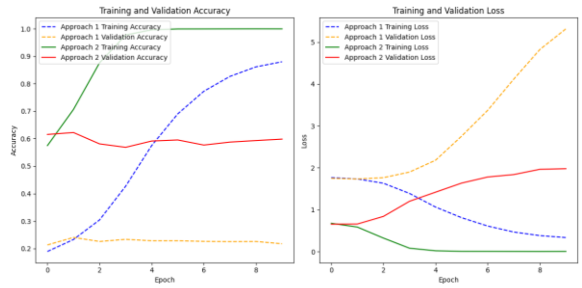
Similarly for the BiLSTM model we see that approach 1 performs worse as compared to approach 2 which is the 2 way approach. Much of the overfitting was reduced in the approach 2 as well as loss was also reduced.

CNN

CNN stands for Convolutional Neural Network. It is a type of deep learning model commonly used for image recognition and computer vision tasks. CNNs are designed to automatically learn and extract relevant features from input images through a series of convolutional layers. These layers apply filters to the input data, capturing different patterns and spatial relationships. CNNs are known for their ability to effectively handle spatial hierarchies and translation invariance, making them highly suitable for tasks such as image classification, object detection, and image segmentation.



Results CNN



Here also for CNN we can see that for both loss and accuracy approach 2 outperforms approach 1.Having lesser overfitting and loss.

Results

	Testing Accuracy		**Validation Accuracy**	
	Classification		**Classification**	
Model	**2-way**	**6-way**	**2-way**	**6-way**
LSTM	56.51	21.39	56.53	21.48
Bi-LSTM	56.59	**22.49**	56.83	**22.86**
CNN	**60.93**	22.10	**59.83**	21.74

Now when comparing all the models three models and 2 approaches which effectively gives us 6 different models to compare, we see that best performance was observed in the case of CNN with 2-way approach. We also see that the 6-way approach doesn't perform that well in any of the cases but Bi-LSTM performed well in that case.

Comparisons from the Research papers

Models	Valid.	Test
Majority	0.204	0.208
SVMs	0.258	0.255
Logistic Regress0ion	0.257	0.247
Bi-LSTMs	0.223	0.233
CNNs	0.260	0.270
Hybrid CNNs		
Text + Subject	0.263	0.235
Text + Speaker	**0.277**	0.248
Text + Job	0.270	0.258
Text + State	0.246	0.256
Text + Party	0.259	0.248
Text + Context	0.251	0.243
Text + History	0.246	0.241
Text + All	0.247	**0.274**

Ref: Liar, Liar Pants on Fire

The first results are from the Liar Liar Pants on Fire research paper published by William Yang. They also used different kinds of models like LSTM, CNN etc. They could manage to get a maximum testing accuracy of 27.4% in the case of CNN, which is very comparable to our case. They just used a 6-way approach considering all the labels.

	2-CLASS		6-CLASS	
	text	+ LIWC	text	+ LIWC
Majority Baseline	.39	-	.06	-
Naive Bayes	.44	.58	.16	.21
MaxEnt	.55	.58	.20	.21
LSTM	.58	.57	.21	.22

Table 5: Model performance on the Politifact validation set.

MODEL	FEATURE	2-CLASS	6-CLASS
Majority Baseline		.39	.06
Naive Bayes	text + LIWC	.56	.17
MaxEnt	text + LIWC	.55	.22
LSTM	text + LIWC	.52	.19
LSTM	text	.56	.20

Table 6: Model performance on the Politifact test set.

Ref: Truth of Varying Shades

Next are the results from Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking by Hannah Rashkin Eunsol Choi Jin Yea Jang Svitlana Volkova Yejin Choi. They majorly used Naive Bayes and basic LSTM for their paper. They were able to get a testing accuracy of maximum 56% for both Naive Bayes and LSTM. Our LSTM also got similar results while our CNN model with 2-way approach outperformed their best models.

Reference

1.William Yang Wang 2017 “Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection https://arxiv.org/pdf/1705.00648.pdf

2.Hannah Rashkin Eunsol Choi Jin Yea Jang Svitlana Volkova Yejin Choi 2017 Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking https://aclanthology.org/D17-1317.pdf

3.Automatic Detection of Fake News 2017 Veronica Perez-Rosas, Bennett Kleinberg , Alexandra Lefevre Rada Mihalcea https://arxiv.org/pdf/1708.07104.pdf

4.Automatic Deception Detection: Methods for Finding Fake News Nadia K. Conroy, Victoria L. Rubin, and Yimin Chen 2015 https://asistdl.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/pra2.2015.145052010082

5.This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News Benjamin D. Horne and Sibel Adalı https://arxiv.org/pdf/1703.09398.pdf

6.A Survey on Fake News and Rumour Detection Techniques Alessandro Bondielli Francesco Marcellonia https://arpi.unipi.it/bitstream/11568/1000990/5/FakeNewsRumoursSurvey.pdf