A Closer Look at Fake News Detection: A Deep Learning Perspective

Ayat Abedalla*
Computer Science Department
Jordan University of Science and Technology
Irbid, Jordan
aaabedalla17@cit.just.edu.jo

Aisha Al-Sadi*
Computer Science Department
Jordan University of Science and Technology
Irbid, Jordan
asalsadi16@cit.just.edu.jo

Malak Abdullah
Computer Science Department
Jordan University of Science and Technology
Irbid, Jordan
mabdaullah@just.edu.jo

ABSTRACT

The increasingly rapid pace of spreading fake news is considered a problem in conjunction with the increasing number of people who are relying upon social media to get news. That earns widespread attention from research communities due to the negative impact and influence of fake news on public decisions. Consequently, the current research strives to illuminate on fake news problem and the process of detecting fake news using deep learning approaches. Using the Fake News Challenge (FNC-1) dataset, we have developed different models to detect fake news based on the relation between article headline and article body. Our models are assembled mainly from Convolutional Neural Network (CNN), Long Short-Term Memory network (LSTM) and Bidirectional LSTM (Bi-LSTM). In the contrary of other studies on the same dataset where they reported accuracy for a test data derived from the same training dataset, our experiments achieved 71.2% accuracy for the official testing dataset.

CCS CONCEPTS

• **Computing methodologies → Machine learning**

Keywords

Fake news; machine learning; deep learning; LSTM; preprocessing; CNN

1. INTRODUCTION

With the rapid growing of data over the internet and the addiction of using social networks in different life aspects, social media become one the main sources of news for many people around the world. This is because of it's rapid spread of information in addition to the low cost of publishing. On the other hand, the bad side effect of spreading news is the appearance of fake news that aim intentionally to mislead people opinions and deceive readers. Fake news are created to influence people's view toward a critical topic or a political agenda, such as in the USA 2016 elections [1]. Moreover, it can also be created to cause a confusion on trading and stock market [2], or to gain a profitable business for online publishers.

The definition of fake news in Wikipedia website: "Fake news, also known as junk news or pseudo-news, is a type of yellow journalism or propaganda that consists of deliberate disinformation or hoaxes spread via traditional news media or online social media"[1]. Due to the difficulty of manually detecting fake news, critical attention has been devoted to detecting fake news and recognizing credible information from several researchers, recently. Although there are different perspectives to study this problem, machine learning methods are considered promising technologies for improving fake news detection solutions. Consequently, a Fake News Challenge (FNC-1) [3] was organized in 2016 to explore how to utilize artificial intelligence techniques, especially machine learning and natural language processing, in combating the problem of fake news. FNC-1 focuses on identifying fake news and stance detection. This paper provides a throughout study on machine and deep learning techniques to crystallize into automatic fake news detection.

The rest of this paper is organized as follows: Section 2 reviews and compares the exiting research works on fake news detection area. Section 3 describes FNC-1 dataset and analyzes it. Section 4 presents our proposed models on detecting fake news using deep learning approaches. While, section 5 provides the results and discussion. Finally, section 6 concludes the paper and presents future work.

2. RELATED WORK

Granik et al. [4] used Naive Bayes (NB) in fake news detection which is a probabilistic model that assumes independence between features. Although this assumption is very simple and naive, it has proved its effectiveness. The researchers collected a dataset consists of 2282 posts from Facebook that shared news articles. In addition to posts' texts, the dataset contains also Facebook engagement information about each post, such as the number of likes, number of comments, number of shares, and information about the content of the post (has a photo, video, or link).

Long et al. [5] proposed a hybrid attention Long Short-Term Memory (LSTM) model. They used LIAR dataset provided by Wang [6], which includes 141 topics in the PolitiFact website[2]. The dataset contains 12,836 news of 3,341

*These authors contributed equally to the work

[1]https://en.wikipedia.org/wiki/Fake-news
[2]http://www.politifact.com/truth-o-meter

ICAAI 2019, October 26-28, 2019, Istanbul, Turkey.

DOI: 10.1145/3369114.3369149

speakers, each news comprises topic, text, and speaker profile. They obtained the representation of news using LSTM and the vector representations of speaker profiles by another LSTM. Then the SoftMax activation classified news based on the two representations. The results showed that their model outperformed modern models that used the reference dataset at 14.5% accuracy. This proved the importance of the speaker's profile in determining the credibility of the news.

Roy et al. [7] worked on the LIAR dataset [6] and used an ensemble model to classify statements into the six classes. The ensemble model consists of two single models which are Bi-directional Long Short-Term Memory (Bi-LSTM) model and Convolutional Neural Network (CNN) model. A multilayer perceptron model has been used to combine these two models.

Thota et al. [8] used a dense neural network (DNN) in their model and made three different experiments based on the word vector representation technique used: BoW, TF-IDF, and Word2ec. A dropout was used as a regularization method to prevent overfitting problem. This work used the dataset provided by Fake News Challenge (FNC-1) that include news articles with their headlines. The data divided into three categories after shuffling randomly, 67% as training and validation data, and 33% as test data. The detection of fake news here is different on the basis that the stance relation between the article and its headline is a key to detect fake news. The accuracies of the models DNN + Word2Vec, DNN + BoW, DNN + TF-IDF are 75.67%, 89.23%, and 94.31% respectively. Table 1 and Table 2 summarize the comparison between the related works.

3. DATASET

3.1 Description and Analysis

In this paper, the Fake News Challenge (FNC-1) 2016 dataset [3] has been used, which includes news articles with their headlines and the stance relation between them. The topics of these articles are varied between political, propaganda, passion, and other real-life topics. The data has been annotated with four class labels based on the relation between the article and its headline that are: agree, disagree, unrelated and discuss. The train data file contains about 50k article headline-body pairs and test data file contains about 25k pairs. Table 3 shows the number of unique headlines and unique article texts for both train and tests data. Its worth mentioning that the train and test data are created from multiple combinations between these unique headlines and bodies. Figure 1 shows the percentage of each label in both train data and test data. Labels distribution is the same in both train and test data exactly. The data is unbalanced, 'unrelated' class is the dominant.

3.2 Data Cleaning and Pre-processing

Data preprocessing is an important task before applying machine learning and deep learning models. For cleaning data, we have lowered all letters, removed stop words, removed punctuations, and stemmed words. We have used NLTK[3] English stop words list for removing stop word. Furthermore, we have removed HTML and replaced the digits with words. We have also unified some abbreviations to be in one form (example I'm to I am). In addition, we have combined each article headline with its body in one text. Table 4 shows the maximum, average, and minimum length of the combined texts (in number of words) in both train and test data.

[3]https://www.nltk.org/

4. PROPOSED MODELS

In this work, we have used:

- Global Vectors for Word Representation (GloVe) [9] that is an unsupervised machine learning algorithm for getting vector representations for words. Each word is mapped to 300-dimension vector to learn the semantics of words.
- Convolutional Neural Networks (CNN) with Max-Pooling: that is considered a deep neural network. The CNNs are able to detect patterns and are used successfully to extract features not only for images but also for text. CNN has two main components: the convolutional layer and the pooling layer. The convolutional layer consists of a set of learnable filters that slide over entire rows of the matrix. After the convolution layer, we use pooling layer to make pattern detection more robust and prevent overfitting data by reducing the complexity and parameters in the network.
- Bidirectional Long Short-Term Memory (Bi-LSTM): a special class of Recurrent Neural Networks (RNN), introduced by Hochreiter and Schmidhuber in 1997 [10]. LSTMs are designed to avoid long-term dependency problem and the gradient vanishing problem.
- Attention Mechanism: that has succeeded in a wide range of tasks such as machine translations [11] and image captioning [12]. We use the attention by computing attention weights to indicate the importance of different words by building a fully connected neural network on top of each encoded state of bidirectional LSTM.
- Dropout Layer: to avoid overfitting.

4.1 Architecture

This subsection is dedicated to describe four proposed models to detect stance between each pair of headline and article body.

- **Model 1 (M1):**

  The first model architecture consists mainly of three CNN layers, followed by one Bi-Lstm layer and one attention layer. Model 1 is summarized in Figure 2.

- **Model 2 (M2):**

  This model has the same architecture of the previous model. However, we have added a batch-normalization and dropout layers. Figure 3 shows the model architecture.

- **Model 3 (M3):**

  This model consists of two LSTM layers that used the hidden states of the first LSTM as the initial states of the second LSTM.

Table 1: Comparison among Fake News Detection research Part 1

Research paper	Data Source	Features Used	Labels
Granik et al. [4]	Facebook posts	Posts texts	Mostly True, Mostly False
Wang et al.[6]	News website, and Kaggle	Text, images, and title	Real, fake
Long et al. [5]	LIAR Dataset	Speaker information, and topic	pants-fire, false, barely-true, half-true, mostly-true, and true
Roy et al.[7]	LIAR Dataset	Texts and speaker information	pants-fire, false, barely-true, half-true, mostly-true, and true
Thota et al.[8]	Fake News Challenge	Article Texts and headlines	agree,disagree, unrelated and discuss

Table 2: Comparison among Fake News Detection research Part 2

Research paper	Main Idea	Model	Metrics	Results
Granik et al.[4]	Classify fake news based on text words	NaÃŕve Bayes	Accuracy	75.4%
Wang et al.[6]	Analysis text and images, extract explicit features, and extract latent features by CNN	TI-CNN	F1-Score	92.1%
Long et al.[5]	Speaker Profiles information	LSTM + Attention	Accuracy	41.5%
Roy et al.[7]	Statement information and Speaker Profile information	CNN + Bi-LSTM	Accuracy	44.9%
Thota et al.[8]	Find the relation between the article and its headline	DNN + TF-IDF	Accuracy	94.31%

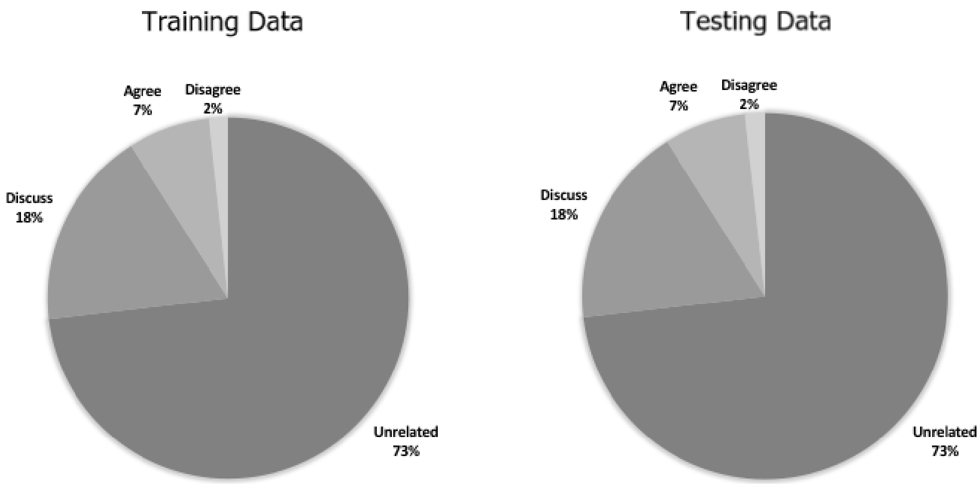

Figure 1: Class Labels Distribution of FNC-1 dataset

Table 3: Number of Unique Headlines and Bodies

	Number of Unique Headlines	Number of Unique Bodies
Train data	1648	1669
Test data	894	900

Table 4: Combined Texts Statistics

	Text MAX Length	Text AVG Length	Text MIN Length
Train Data	2842	225.78	6
Test Data	1944	212.83	6

- **BERT Model:**

  Transformer model proposed in 2017 [13], which is based on attention mechanism and achieved the state of the art in wide range of NLP tasks. in 2018, Google proposed a new model called Bidirectional Encoder Representations from Transformers (BERT) [14]. BERT is a new language model can be used in NLP tasks such as text classification. To the best of our knowledge, none has used BERT with FNC dataset. We have used the BERT model which gave a competitive accuracy that will be discussed in details in the next section.

4.2 Experiments and Results

We have made different experiments for each model for optimization purposes. Starting with the first model (M1), we have carried three experiments (M1-E1, M1-E2, and M1-E3) with different hyper-parameters values for each of the used optimizer, learning rate, batch size, and number of training epochs. more details are as follows:

- Experiment 1 (M1-E1): In this experiment, we have split data into 70% train and 30% validation. Each of the three CNN layers is followed by a max pooling layer, then the result has been fed into Bi-LSTM with applying the attention technique to enable the network to focus on relevant parts of the input. This model achieves accuracy as follows 90% for validation set, 55.21% for competition test set, and 47.67% score

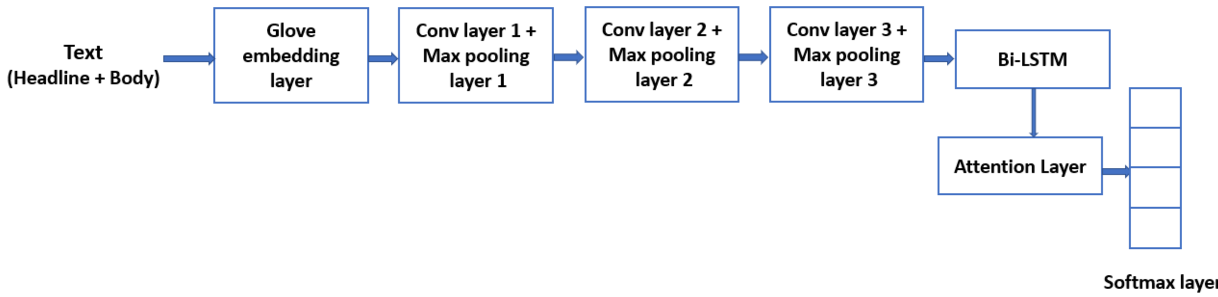

Figure 2: First Proposed Model (M1)

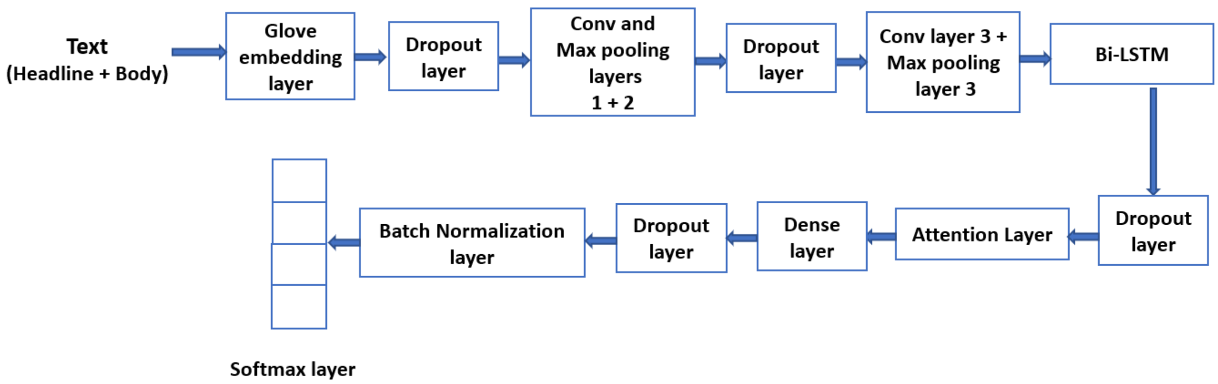

Figure 3: Second Proposed Model (M2)

for competition test set.

- Experiment 2 (M1-E2): The settings of this experiment are similar to E1. However, after the second layer of CNN, we have added a dropout layer. Then a Bi-LSTM with 300 units is used before applying the attention technique. At the end, we have used the softmax layer. This model achieves accuracy as follows 86.35 for validation set, 72.20% for competition test set, and 45.497% score for competition test set.

- Experiment 3 (M1-E3): In this experiment, we split data into 70% train, 15% validation, and 15% test. In this experiment, we have added a dropout layer after the embedding layer to mask some words and learn to make correct predictions. Then we have added three CNN layers, the first two layers contain 128 hidden units, and the third layer contains 256 units. Then add Bi-LSTM followed by (0.1) dropout and attention layer followed by (0.1) dropout. Finally, we have used the softmax layer. This model achieves accuracy as follows 87.47% for validation set, 87.89% for test-train set, 55.31% for competition test set, and 50.97% score for competition test set.

For the other two models, we have split data into 70% train and 30% validation. The second model experiment (M2-E) achieves accuracy as follows 93.4% for validation set and 71.21% for competition test set, and 57.52% score for competition test set. While the third model experiment (M3-E) achieves accuracy as follows 73% for validation set, 73% for competition test set. But its misleading because it predicts 'unrelated' class all the time. Finally, for BERT model, its worth mentioning to say that We have used the smallest BERT model due to the memory resources. The training process took several hours and the test accuracy was 18%, which is very low compared to other models. This is due to resources limitations, therefor the work on this model has been recorded as a future work. Table 5 summarizes the different experiments parameters settings. While all experiments results are summarized in Table 6.

Table 5: Experiments Settings

Experiment	Batch size	Optimizer	Learning Rate	Epochs
M1-E1	80	RMSprop	0.001	30
M1-E2	64	Adam	0.001	9
M1-E3	64	RMSprop	0.001	4
M2-E	64	Adam	0.001	50
M3-E	32	Adam	0.0001	3
BERT-E	16	Adam	2e-5	25

Table 6: Experiments Results

Experiment	Val-Acc	Test-Train-Acc	Test-Acc	Test-Score
M1-E1	90	N/A	55.21	47.67
M1-E2	86.35	N/A	72.20	45.497
M1-E3	87.47	87.89	55.31	50.97
M2-E	93.4	N/A	71.21	57.52
M3-E	73	N/A	73	N/A
BERT-E	97.8	N/A	18	N/A

5. DISCUSSION AND EVALUATION

In the original Fake news challenge, the evaluation method gave different weights for predicting labels as follows:

- 25% to the score weight given for predicting the label if related or unrelated.
- 75% to the score weight given for predicting the specific related label (Agree, Disagree, and Discusses).

For evaluation purposes and to compare our result with other researcher's experiments results, we have realized that most of the existing work used a traditional accuracy measure with equal weights for all labels, which is less strict than the original score evaluation used by the competition. We have also realized that the existing work had ignored completely the test data and used a random split form the training data set to evaluate the models. Knowing the fact that the train dataset contains many repeated articles (headlines and bodies), while the official test data does not have any intersection with train data. Therefore, taking a random records from training data to be considered as a test data will give a higher accuracy. Table 2 shows the different accuracies.

In our work, we have adopted the original test data to compare the results using two evaluation method as shown in Table 6. Our best result is 71.2% with the traditional equal-weights scoring and 57.52% with the original scoring provided by the competition website. We was not able to compare our results with other existing work since there is no standard dataset with their work.

6. CONCLUSION AND FUTURE WORK

Nowadays, people are relying upon social media to read the news instead of traditional news media. This leads to a massive spread of fake news due to the low quality of online news compared with traditional news from official organization. The large volume of fake news with intentionally false information to mislead the reader has a negative impacts on society. In this research, we investigated the fake news problem by reviewing the existing work for detecting fake news using machine learning and deep learning fields. In addition, we built four deep learning models using CNN and LSTM networks to detect fake news based on the relation between article headline and body of FNC-1 dataset. The best proposed model achieved 71.2% accuracy on the official test data that. Up to our knowledge, this score has not been achieved in previous studies. As a future work, we aim to improve the models by getting more insight on the data. Also, we strongly believe that using BERT model would have a strong impact on the results.

7. REFERENCES

[1] Alexandre Bovet and HernÃạn A. Makse. Influence of fake news in twitter during the 2016 us presidential election. *Nature Communications*, 10(1), 2019.

[2] K Rapoza. Can âăÿfake newsâăź impact the stock market. 2017.

[3] Fake News Challenge. Fakenewschallenge.org. http://www.fakenewschallenge.org/, 2019. [Online; accessed 28-April-2019].

[4] Mykhailo Granik and Volodymyr Mesyura. Fake news detection using naive bayes classifier. *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 2017.

[5] Yunfei Long, Qin Lu, Rong Xiang, Minglei Li, and Chu-Ren Huang. Fake news detection through multi-perspective speaker profiles. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, volume 2, pages 252–256, 2017.

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

[7] Arjun Roy, Kingshuk Basak, Asif Ekbal, and Pushpak Bhattacharyya. A deep ensemble framework for fake news detection and classification. *arXiv preprint arXiv:1811.04670*, 2018.

[8] Aswini Thota, Priyanka Tilak, Simrat Ahluwalia, and Nibrat Lohia. Fake news detection: A deep learning approach. *SMU Data Science Review*, 1(3):10, 2018.

[9] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[10] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[11] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[12] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057, 2015.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.