# Fake News Detection(Group: 19)

**Instructor - Prof. Harish Karnick**

**Praphul Singh** 150513
**Robin Singh** 150591
**Harshit Saini (Auditor)** 150279
**Tushar Gupta** 150770
**Asitava Sarkar** 150160
**Shashwat Singh Jamwal** 150667
**Ribhu Nirek** 150578

## Abstract

The challenge to combat the menace of fake news is essential for the maintenance of authenticity of various media sources. We have used machine learning and natural language processing to identify the fake news which can be used to combat the fake news problem. We have used two models, Attention Integrate(AI) model and 3HAN (Hierarchical Attention Network) **[SFR17]** in our project. We have focused on using the attention mechanism to check the relevancy of words and sentences in the articles. We merged the context info from both headlines and articles to decide the genuineness of the article.

## 1 Introduction

Fake news are made up stories with the intention to deceive, often with monetary gain as a motive. It is a type of yellow journalism or propaganda that consists of purposeful misinformation. It undermines serious media coverage and makes it more difficult for journalists to cover significant news stories. The general sentiment of the article is provocative and has a typical headline and writing style. It is typically generated for commercial interests like attracting viewers and collecting advertising revenue. It is believed that fake news also had a large impact on the recent US Presidential elections. Many companies like Google and Facebook have put hard efforts to classify these kind of malicious information.

Google has implemented a page ranking algorithm which gives more preference to the genuine news and has banned the sources which have been recorded in spreading fake news. We are all aware of use of social media platforms like facebook, twitter used to run the propaganda more generally politically motivated. There is also some work done to identify the fake news spread through twitter where large no. of hired people for that purpose and nowadays bots are also used for the same. The misinformation is also spread using the doctored videos and images which is difficult to identify as it is also a kind of fake news but with no text content or very less content. Hence, in future we expect that there will also be some method evolving to identify such doctored videos and images.

## 2 Previous Work

**Fake news detection has been a widely studied topic in recent times. A variety of techniques have been proposed in this regard to solve this problem. This section briefs about some techniques which have been successful in solving this problem up to a certain extent.**

### 2.1 A Survey on Fake News Detection, Stanford University(Winter 2017)

This project is a survey on various methods that can be implemented to solve the fake news detection problem. The dataset used for the project was the kaggle dataset for fake news and signal media news dataset for the authentic news. Glove embeddings of dimension 300x1 were used to represent the words as vectors.

#### 2.1.1 Logistic Regression

First model was a simple logistic regression based model in which the news article vector was created by taking the average of the individual word embeddings and then prediction was made by using a softmax layer.

#### 2.1.2 Feed Forward Neural Network

The second method focuses on a two layered feed forward neural network in which the input was the averaged vector of the word embeddings and the output was a single digit 0 or 1 depicting whether the article is fake or not. For non-linear activation, Relu Function was used and to reduce overfitting a dropout layer was also used.

#### 2.1.3 Recurrent Neural Networks

RNN was used to represent the article as a single vector. But as these layers cannot memorize infromations for long sequences LSTMs and GRUs were used to represent the text as vectors and finally softmax was applied to predict the label.

**Other approches like CNN and attention layers were also used to improve and experiment the performances of the models. The statistics below is a comparison of their accuracies:**

| Model | Precision | Recall | F1 Score |
|---|---|---|---|
| Logistic Regression | 0.96 | 0.49 | 0.65 |
| Feed Forward Neural Network | 0.89 | 0.74 | 0.80 |
| RNN | 0.91 | 0.56 | 0.70 |
| GRUs | 0.89 | 0.79 | 0.84 |
| LSTMs | 0.93 | 0.72 | 0.81 |
| BiLSTMs | 0.88 | 0.75 | 0.81 |
| CNN | 0.87 | 0.44 | 0.58 |
| CNN with Attention | 0.97 | 0.03 | 0.06 |

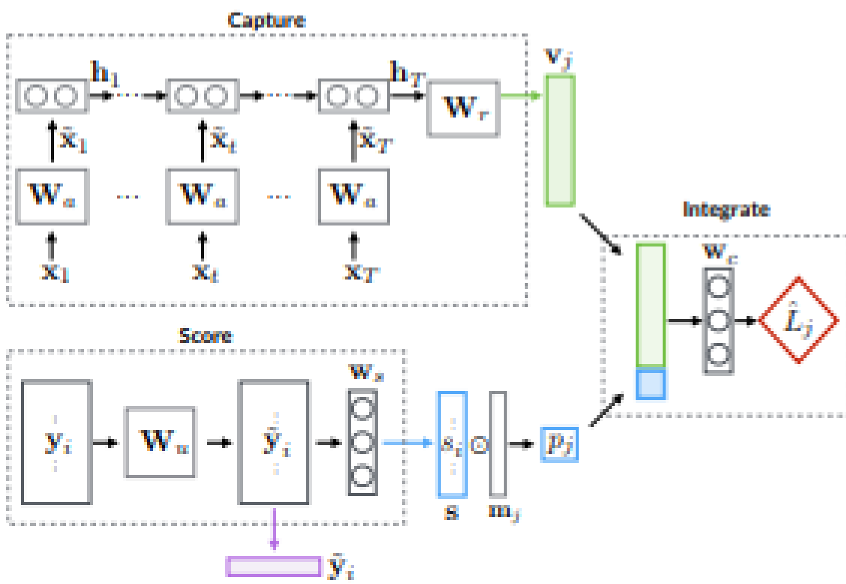
### 2.2 CSI: A Hybrid Deep Model for Fake News Detection [RSL17]

#### 2.2.1 Introduction

This paper uses three characteristics of any news article to predict whether the article is fake or not. The first characteristic was the text of the article which evaluates whether the headline matches with the body of the article, judging the consistency and the quality of the language. The second characteristic was the response that a news article was meant to illicit examining whether the article contains opinionated and inflammatory language, crafted as a bait or to incite confusion. The third characteristic used was source of the article which checks the structure of the url, credibility of the media source, profile of the journalist who authored it.

#### 2.2.2 Model

The CSI(Capture, Score and Integrate) model was trained on two social media datasets TWITTER and WEIBO separately and then tested on them. It consists of three parts namely Capture, Score and Integrate. Capture part was used to extract temporal representations of articles using Recurrent Neural Network(RNN). Temporal engagements were then stored as vectors and were fed into the RNN which produces an output, a representation vector. After that, Score part was used to compute a score and representation. The user-features were fed into a fully connected layer and a weight was applied to produce the scores vectors. Finally, Integrate part was used to concatenate the outputs of the two modules and the resultant vector was used for classification.




**C**apture **S**core **I**ntegrate Model Architecture

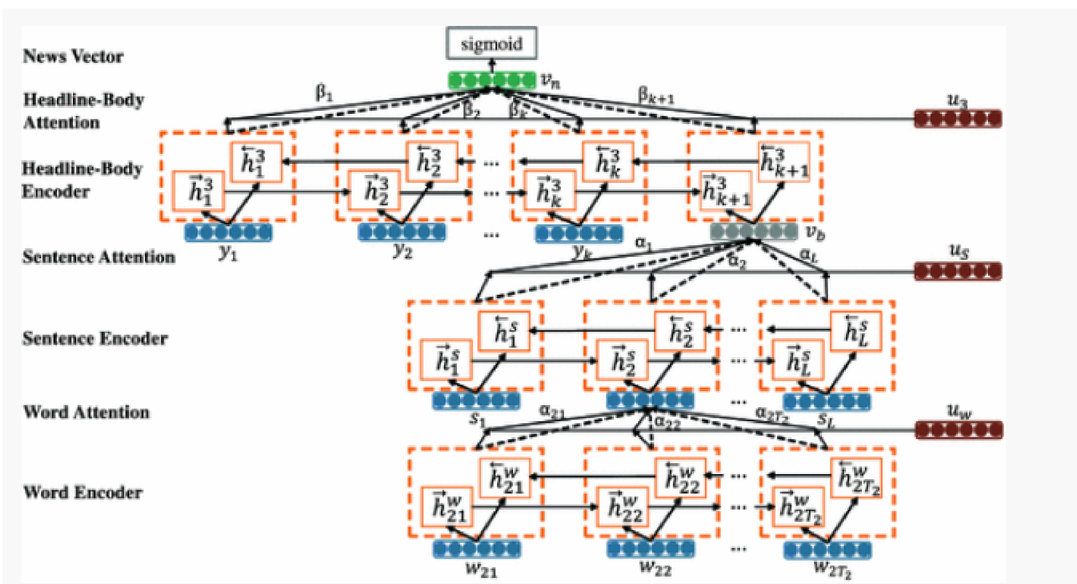
### 2.3 3HAN: A Deep Neural Network for Fake News Detection [RSL17]

#### 2.3.1 Introduction

This paper presents a technique which in authors' words is a three level Hierarchical Attention Network (3HAN) which creates an effective representation of a news article called news vector. The news vector can be used to classify an article by assigning a probability of being fake. 3HAN provides an importance score for each word and sentence of an input article based on its relevance in arriving at the output probability of that article being fake. These importance scores can be visualized through a heatmap, providing key words and sentences to be investigated by human fact-checkers.

#### 2.3.2 Model

In 3HAN, the structure of an article is interpreted as a three level hierarchy modelling article semantics on the principle of compositionality. Words form sentences, sentences form the body and the headline with the body forms the article. An effective representation of an article is hypothesized using the hierarchy and the interactions between its parts. These interactions take the form of context of a word in its neighbouring words, coherence of a sentence with its neighbouring sentences and stance of a headline with respect to the body. Words, sentences and headline are differentially informative dependent on their interactions in the formation of a news vector. Finally these three layers of attention mechanisms are combined to exploit this differential relevance.




3HAN Model Architecture

## 3 Dataset*(s)*

The data that we propose to use in the project is drawn from three different resources. We are extracting both fake and genuine news articles from different sources and joining them randomly. Fake news articles were taken from publicly available *Kaggle Dataset* [1]– containing 13,000 fake articles each having different attributes like title, text, source, country where it is published etc.The genuine news articles were extracted from *Signal Media News Dataset* [1]– 40,000 articles were selected at random from the dataset. Therefore, around 13,000 fake and 20,000 genuine are taken for the analysis.

The dataset was divided into 80:20 ratio for training and testing. The dataset mainly contained news articles from US media sources, this leads to an inherent bias in the dataset. As a result, our model performed well on news reports from US media and not well for news from other parts of the world.

## 4 Data Preprocessing

The unnecessary symbols were removed leaving behind alphanumeric characters, stop symbols and apostrophe. Then abbreviations such as haven't, don't etc. were replaced with have not and do not respectively. The sentences were tokenized into words which were then used in our model. The title, text and label were made into 1d array , 2d array of words and a float value respectively. The 2d array of words basically means that the sentences were stored as rows and words of that sentence were stored in the columns. The words were replaced by their 100d glove **[PSM14]** embeddings. This data was then directly fed into the model.

Our Code is available at `https://github.com/robsr/fake-news-detection`

## 5 Our Model

### 5.1 Attention Integrate Model

In this project, we propose a model inspired from the CSI model mentioned above. The difference is that we instead of taking the source as a feature vector, use the headline as the feature vector. We took inspiration from 3HAN to use attention mechanism to focus on the relevant words in the articles.

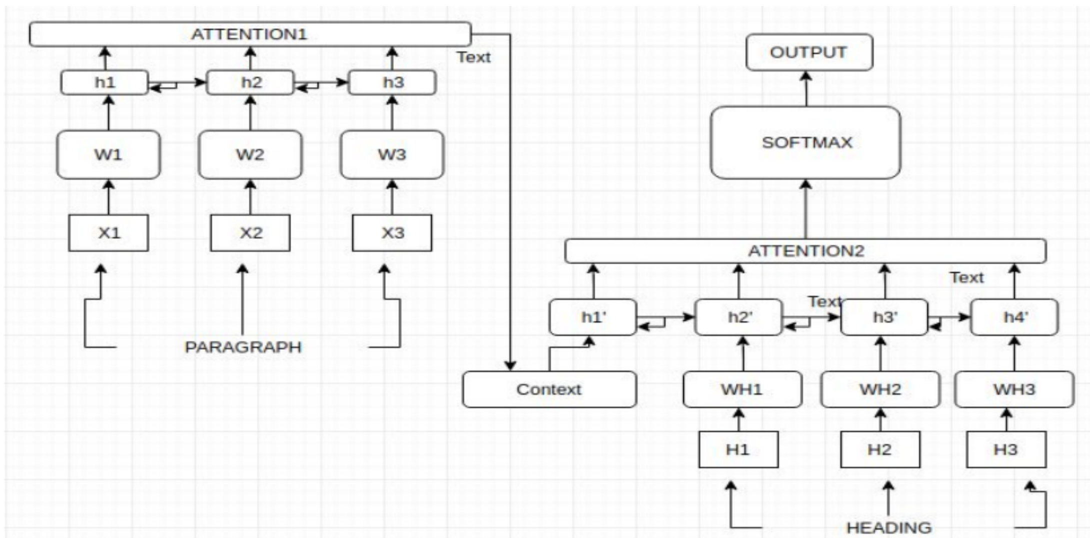
#### 5.1.1 Attention Mechanism [Xu+15]

Attention can simply be understood as a weight vector learnt by the model to focus on certain features. Features can be words or sentences both in our case! **Learning attention weights-**

$$\mathbf{u} = \omega^T tanh(H)$$
$$\mathbf{a} = softmax(\mathbf{u})$$
$$\mathbf{c} = H\mathbf{a}^T$$

The architecture of the AI model is the figure given below. The paragraph is basically a sequence of words here which is converted into a sequence of embedding vectors. The word embeddings used here are taken from Glove vectors of 100x1 dimensions. The embeddings are fed into the bidirectional LSTM layer consisting of 50 hidden units. The output from the LSTM layers gives us a sequence of vectors each of which represents some meaning around the context around the corresponding word. LSTM gate equations with parameters.

$$
\begin{gathered}
f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \\
i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\
o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \\
c_t = f_t o c_{t-1} + i_t o \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \\
h_t = o_t o \sigma_h(c_t)
\end{gathered}
$$

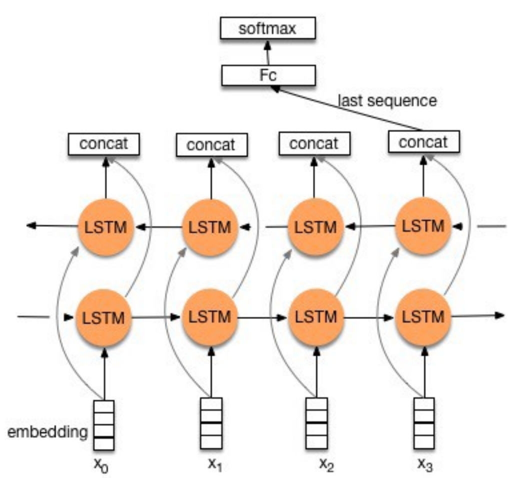



**AI** Model Architecture

The output sequence is then fed into the attention1 layer which learns a weight vector. This is done basically to pick out the relevant meanings from the input sequence. The output is a context vector.

The context is a 100x1 dimensional vector which is then concatenated along with the headline word embeddings to form a new sequence. The purpose of this step is to have a simultaneous understanding of the article along with its headline.

The concatenated sequence is then again fed in a bidirectional LSTM layer which again gives us a sequence of contextual meanings in the form of a sequence of float vectors.




Bidirectional LSTM

The output is then fed into the second attention layer. This is done to check the relevancy between article and headline. The output of the attention layer2 is finally fed into a softmax layer to predict the output.

## 6 Results

3HAN model was trained for 9 epochs and AI model was trained for 15 epochs. The rest results were taken from the survey done by the Standford team and original authors of 3HAN.

| | Twitter | |
|---|---|---|
| | Accuracy | F-score |
| DT-Rank | 0.624 | 0.636 |
| DTC | 0.711 | 0.702 |
| SVM-TS | 0.767 | 0.773 |
| LSTM-1 | 0.814 | 0.808 |
| GRU-2 | 0.835 | 0.830 |

**Twitter Sentiment Analysis**

| Model | Accuracy | F1 Score |
|---|---|---|
| 3HAN | 0.876 | 0.872 |
| AI | 0.862 | 0.858 |

| Model | Precision | Recall | F1 Score |
|---|---|---|---|
| Logistic Regression | 0.96 | 0.49 | 0.65 |
| Feed Forward Neural Network | 0.89 | 0.74 | 0.80 |
| RNN | 0.91 | 0.56 | 0.70 |
| GRUs | 0.89 | 0.79 | 0.84 |
| LSTMs | 0.93 | 0.72 | 0.81 |
| BiLSTMs | 0.88 | 0.75 | 0.81 |
| CNN | 0.87 | 0.44 | 0.58 |
| CNN with Attention | 0.97 | 0.03 | 0.06 |

## 7 Future Work

As mentioned in the dataset section, the dataset mainly consisted of news items from US media sources. An immediate extension of the model can be training the model on a wider variety of news articles covering various fields, demographics  countries, so that it is not just limited to US  politics.

In the recent years, a lot of Silicon Valley companies especially Facebook, Twitter etc. have acknowledged the huge problem posed by fake news on the interest and its possible consequences. They have started to actively work on finding a powerful and practical solution to this problem. A possible approach for this can be analyzing how the fake news spreads on the social media and this data can be used to identify users and media sources which contribute to this menace and blacklist them.

The writing style of the fake news articles have changed significantly and of lately it is becoming increasingly difficult to identify fake news and a significant percent of people are not able to distinguish fake from true news. A mechanism can be invented which also verifies the information and claims of an article from other reputed sources and takes that into account.

Sometimes, distinguishing an article as fake and not fake is not sufficient. It is essential to know which part of the article actually contains the news which is fake. With advancements in this field in the future, this can be hopefully accomplished.

## References


[PSM14] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. "GloVe: Global Vectors for Word Representation". In: *Empirical Methods in Natural Language Processing (EMNLP)*. 2014, pp. 1532–1543. URL: http://www.aclweb.org/anthology/D14-1162.

[Xu+15] Kelvin Xu et al. "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention". In: *CoRR* abs/1502.03044 (2015). arXiv: 1502.03044. URL: http://arxiv.org/abs/1502.03044.

[RSL17] Natali Ruchansky, Sungyong Seo, and Yan Liu. "CSI: A Hybrid Deep Model for Fake News". In: *CoRR* abs/1703.06959 (2017). arXiv: 1703.06959. URL: http://arxiv.org/abs/1703.06959.

[SFR17] Sneha Singhania, Nigel Fernandez, and Shrisha Rao. "3HAN: A Deep Neural Network for Fake News Detection". In: *Neural Information Processing*. Ed. by Derong Liu et al. Cham: Springer International Publishing, 2017, pp. 572–581. ISBN: 978-3-319-70096-0.


## CONTRIBUTIONS

| Name | Percentage |
|---|---|
| Praphul Singh | 27% |
| Robin Singh Rana | 26% |
| Tushar Gupta | 12% |
| Asitava Sarkar | 12% |
| Harshit Saini(Auditor) | 10% |
| Shashwat Singh Jamwal | 8% |
| Ribhu Nirek | 5% |