UCI University of California, Irvine

Deep Networks for Stance Detection (Fake News Challenge-1)

Gagan Khanijau
34166010

Harsha Nadimpalli
63249583

Anirudh Rachuri
47365112

Avinash Aita
88289773

What is ‘Fake news’?

- Fake news is defined as “**a made-up story with an intention to deceive**”

Why is it important?

- These made-up stories can create a great deal of confusion in the public about the facts of current events, thus poses a serious challenge to the news industry as well as the whole of society.

Fake News Challenge

- Goal: Explore how artificial intelligence technologies, particularly machine learning and NLP, might be leveraged to combat the fake news.

- Benefit: Automating parts of the procedure human fact checkers use today to determine if a story is real or a hoax.

Stance Detection Problem

Input

- A headline and a body text - either from the same news article or from two different articles.

Output

- Classify the stance of the body text relative to the claim made in the headline into one of four categories:
  - Unrelated
  - Agrees
  - Disagrees
  - Discusses

Unrelated

- **Zombie Cat: Buried Kitty Believed Dead, Meows Back to Life'**

Hewlett-Packard is officially splitting in two. Following rumors over the weekend, HP is announcing today that it will separate its PC and printer division from its enterprise and services business. The split means current CEO Meg Whitman will become the chairman of the PC and printer operation, and continue as CEO of the split-off enterprise business.

Agree

- **Man saved from bear attack - thanks to his Justin Bieber ringtone**

Justin Bieber may not have been able to take on Orlando Bloom, but he sure as hell was able to take on a bear.

No, PETA, Bieber didn’t beat down a bear, he just scared one away in Russia's Yakutia Republic.

It all went down when 42-year-old fisherman Igor Vorozhbitsyn’s attack by a brown bear was interrupted by a Justin Bieber ringtone.

Disagree

- **Kushner Proposed Backchannel Communication with Russians**

The idea of a permanent back channel was never discussed, according to the source. Instead, only a one-off for a call about Syria was raised in the conversation.

In addition, the source told Fox News the December meeting focused on Russia's contention that the Obama administration's policy on Syria was deeply flawed.

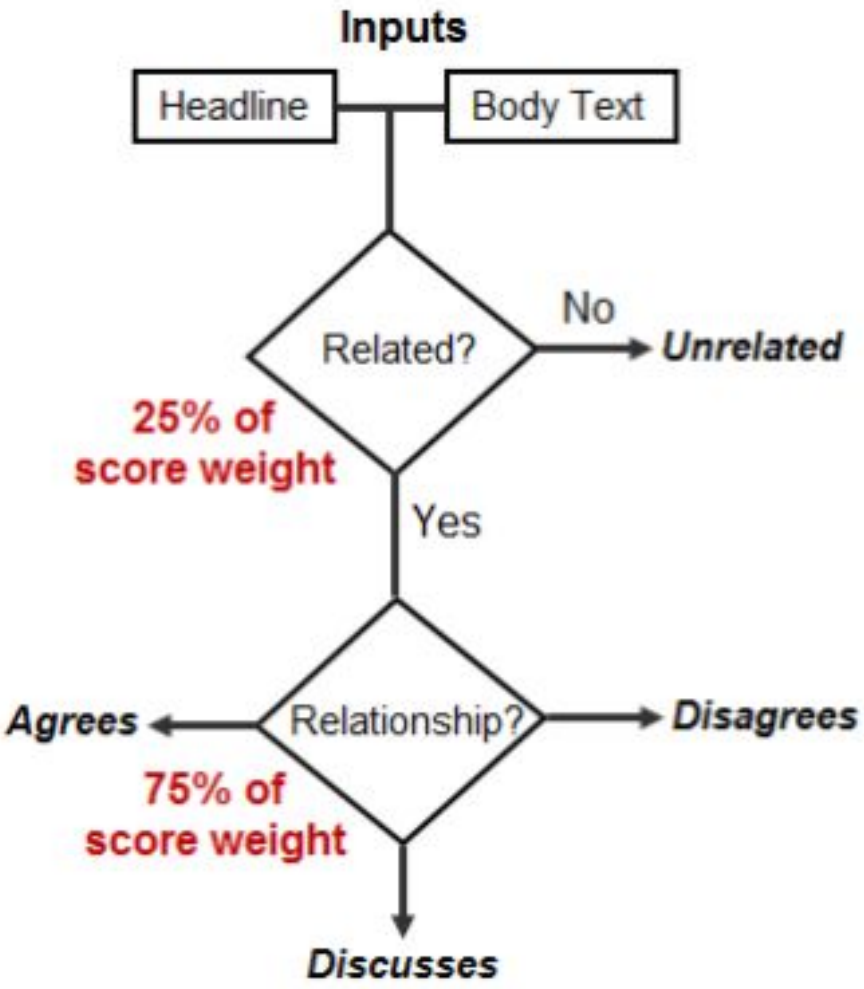
Discuss

- **Steven Sotloff 'beheading': Islamic State fighters release video showing death of second US journalist**

The Islamic State of Syria and Iraq has reportedly released a video showing the beheading of Steven Joel Sotloff, a U.S. journalist being held by the group.

The New York Times, Reuters and the Washington Post cited a report from the SITE Intelligence Group on the video that purportedly showed Sotloff's killing.

Scoring


Inputs
Headline
Body Text
Related?
No
Unrelated
25% of
score weight
Yes
Relationship?
Agrees
Disagrees
75% of
score weight
Discusses

Dataset

- The data is derived from the **Emergent Dataset** created by Craig Silverman.
- Training data size: 49972 pairs of headline and body

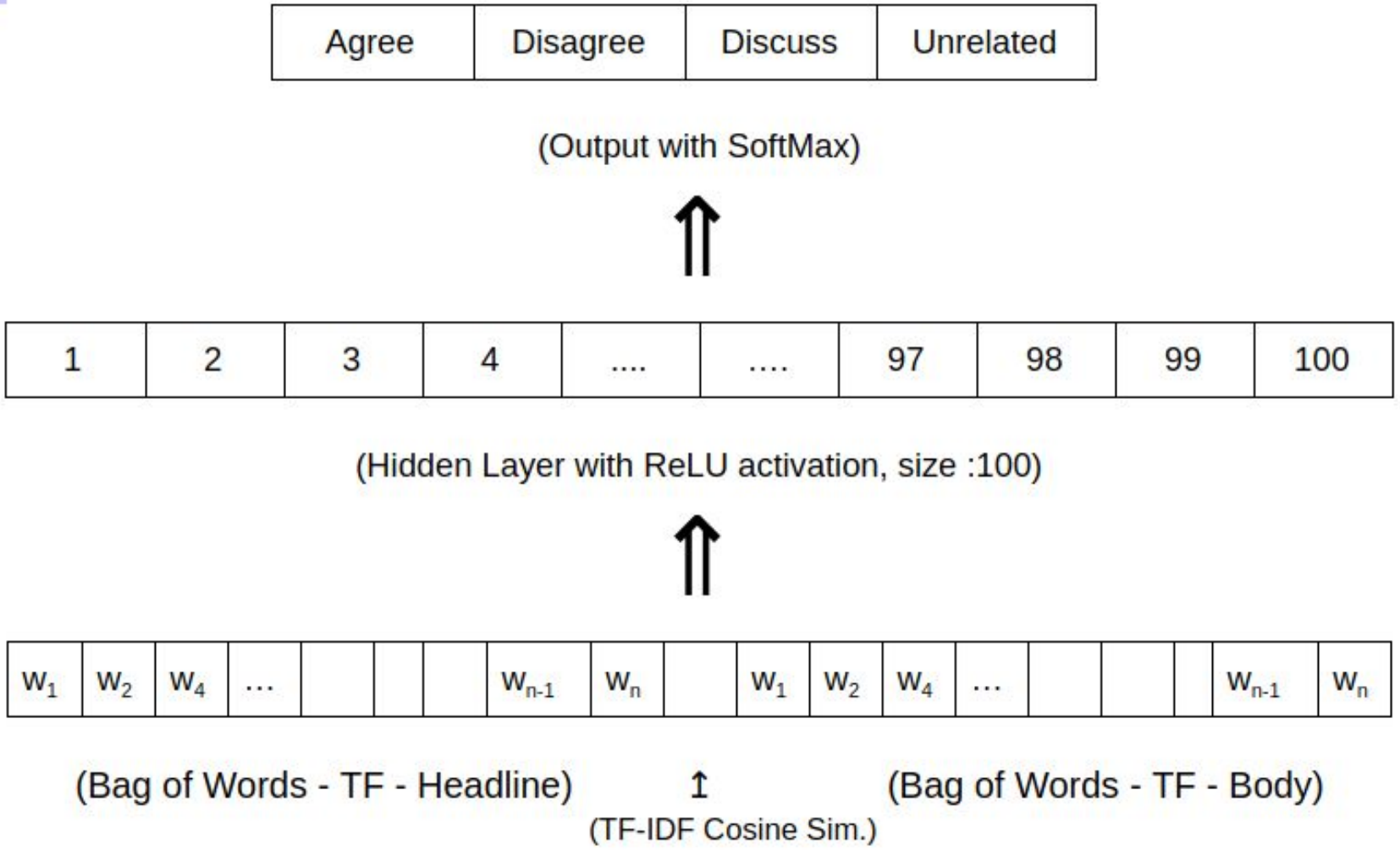
%	unrelated	discuss	agree	disagree
49972	0.73131	0.17828	0.073601	0.016809

- Test data size: 25419 pairs of headline and body

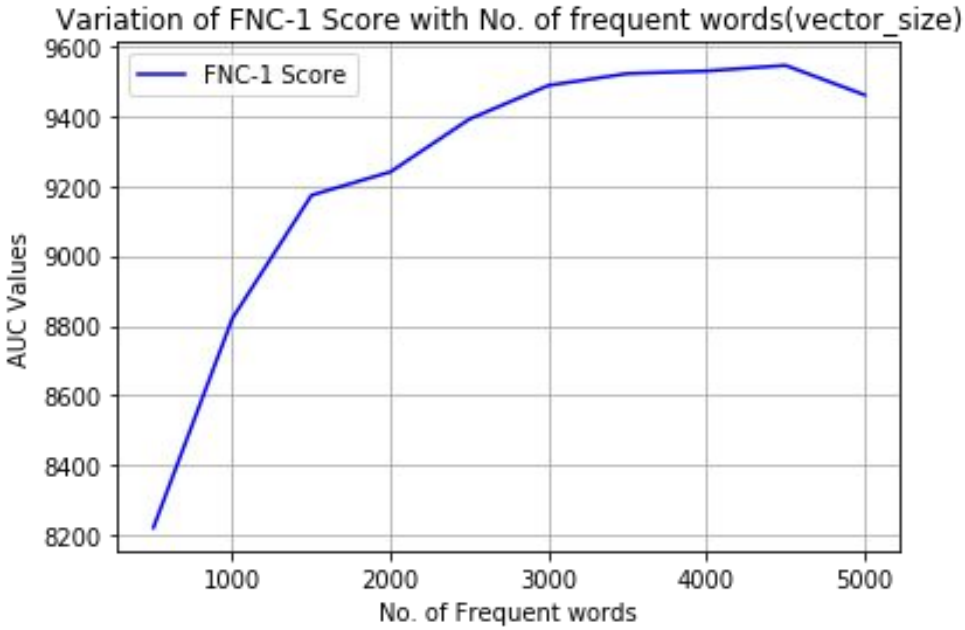
Feed Forward Networks

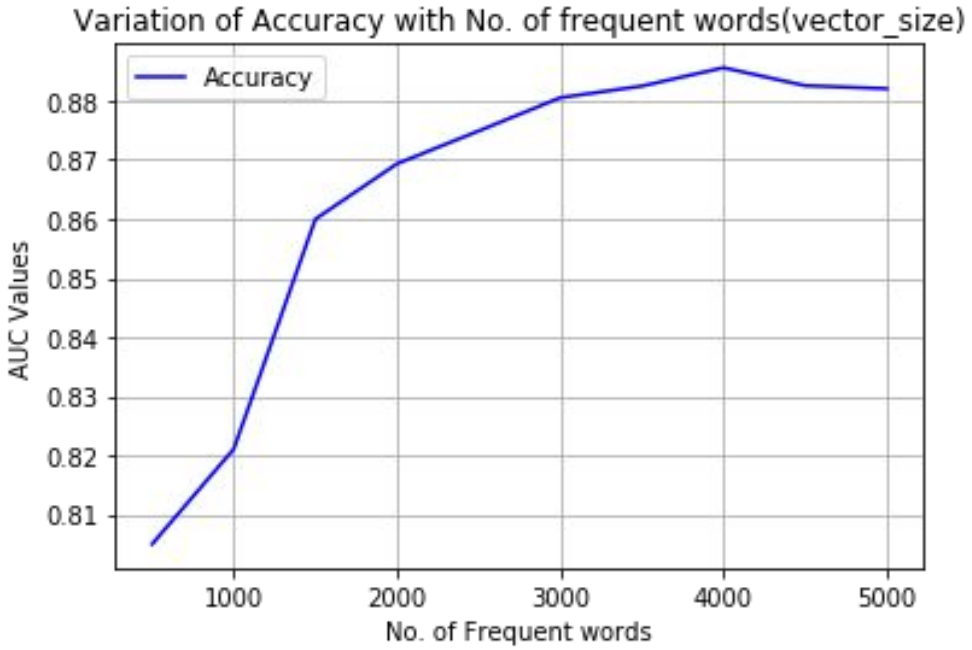
- A Multi - Layered Perceptron model - single Hidden layer with ReLU Activation
  - No. of nodes : 100 ; Hidden Layers : 1 (currently)
- Input : Bag of words term frequency(TF) vector created using most frequent words.
- It is used to predict one of the 4 classes as the stance for the pair of news headline and body.
- Other added feature: Also included cosine similarity between the pair as an extra input feature.

Model


Agree
Disagree
Discuss
Unrelated
(Output with SoftMax)
1
2
3
4
....
....
97
98
99
100
(Hidden Layer with ReLU activation, size :100)
w1 w2 w4 … wn-1 wn w1 w2 w4 … wn-1 wn
(Bag of Words - TF - Headline)
(Bag of Words - TF - Body)
(TF-IDF Cosine Sim.)

Performance





Current Model

- Input : Experimented with different vector size.
- Output (O) = 4
- Hidden (h_i)= 100
- Regularizer (L2): α= 0.00001
- Learning rate = 0.01
- epochs = 90
- Batch size= 500

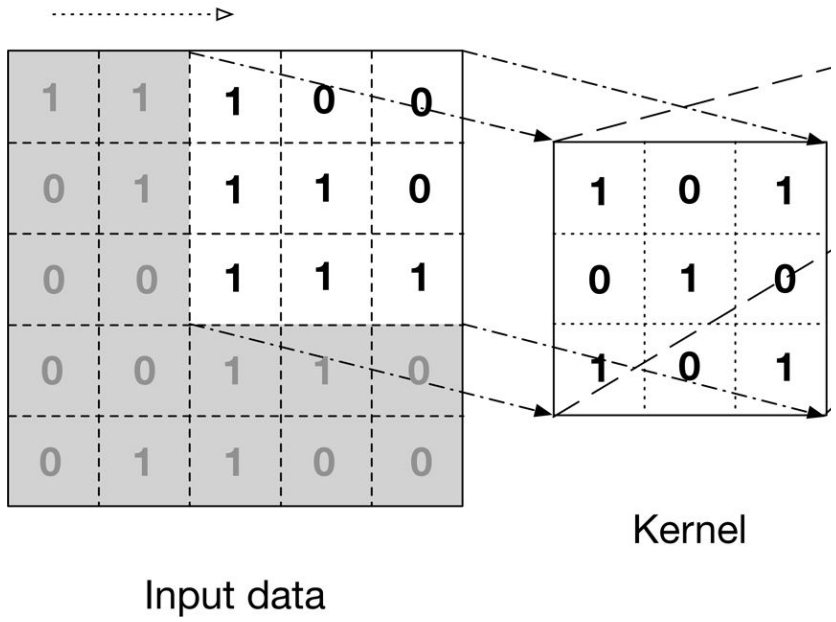
Further Improvements

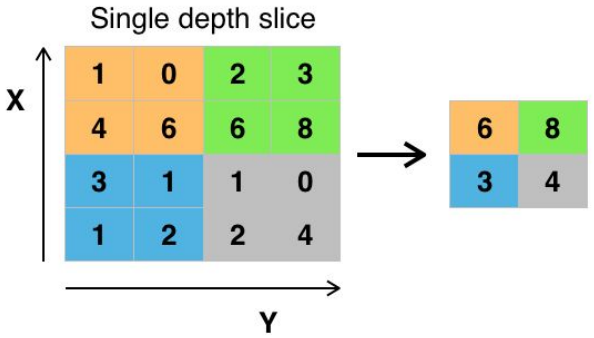
- Can add new features like: n-grams, POS & NER tagged features.
- Variation of hyperparameters in the neural network.
- Variation in number of nodes and layers

Convolutional Neural Network

- **Features:**
  - Word - Vector representation
  - Convolution
  - Spatial Vs Temporal Pooling

2D Visualization

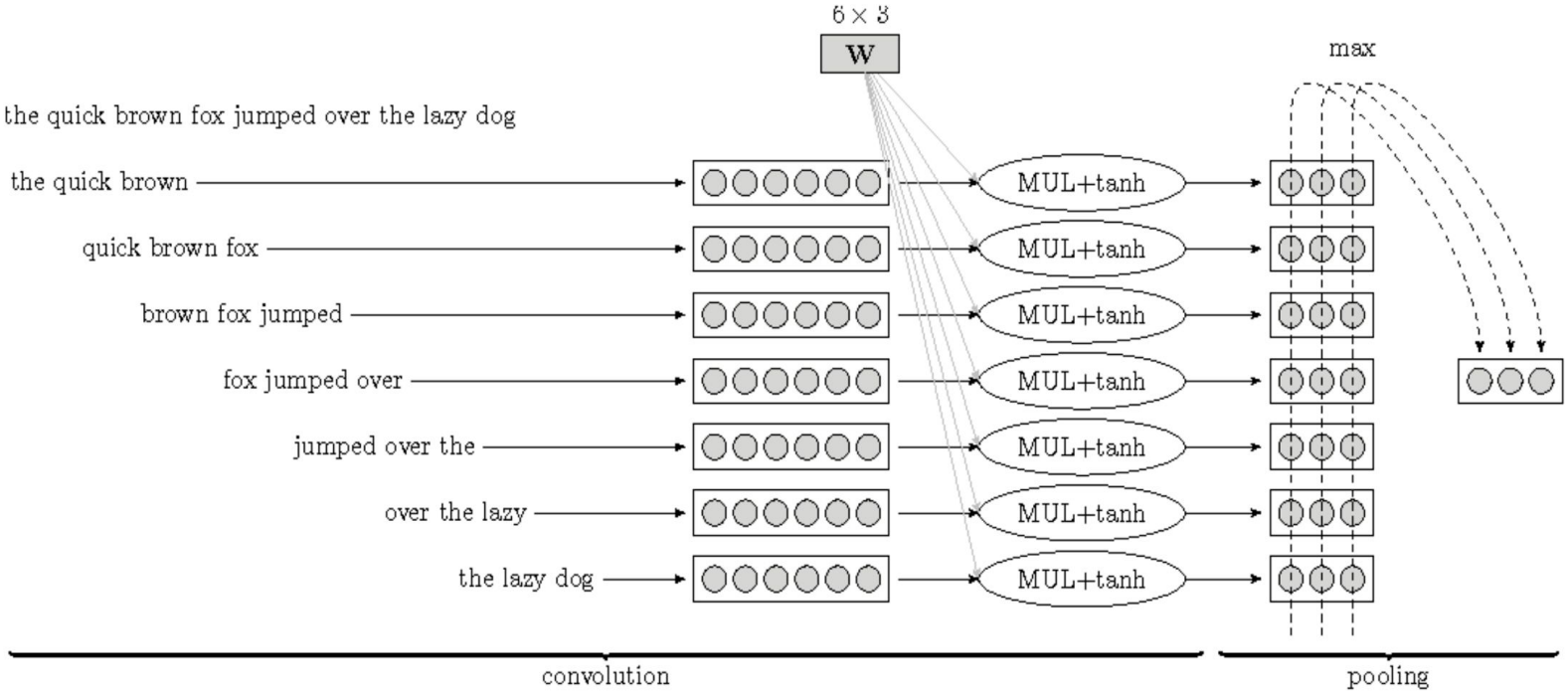




Convolutional Neural Network

- **1D Visualization:**

Interesting Properties: N-Grams, Representative words?



Note: Our network does not replicate this architecture. This image is for intuition and future work idea!

Convolutional Neural Network

- **Survey**:
  - **General Usage:**
    - Images
  - **Speech/Text:**
    - Conversational Speech Recognition - Microsoft
    - Text Summarizers
    - Auto Encoders
- **Motivation**:
  - Stance Detection ≡ Text Summarizer ?
  - Convolution ≡ Auto Encoder / N-gram model?
  - 1D Temporal Pooling ≡ Representative word?
  - 1D Temporal Max Pooling ≡ Encoded Message?
  - Advantage:
    - GPU - Parallel processing
    - Quick learning
- **Important Visaulization:**
  - Word Sequence is considered Temporal

Convolutional Neural Network

Architecture:

Hyper Parameters: Not Optimized

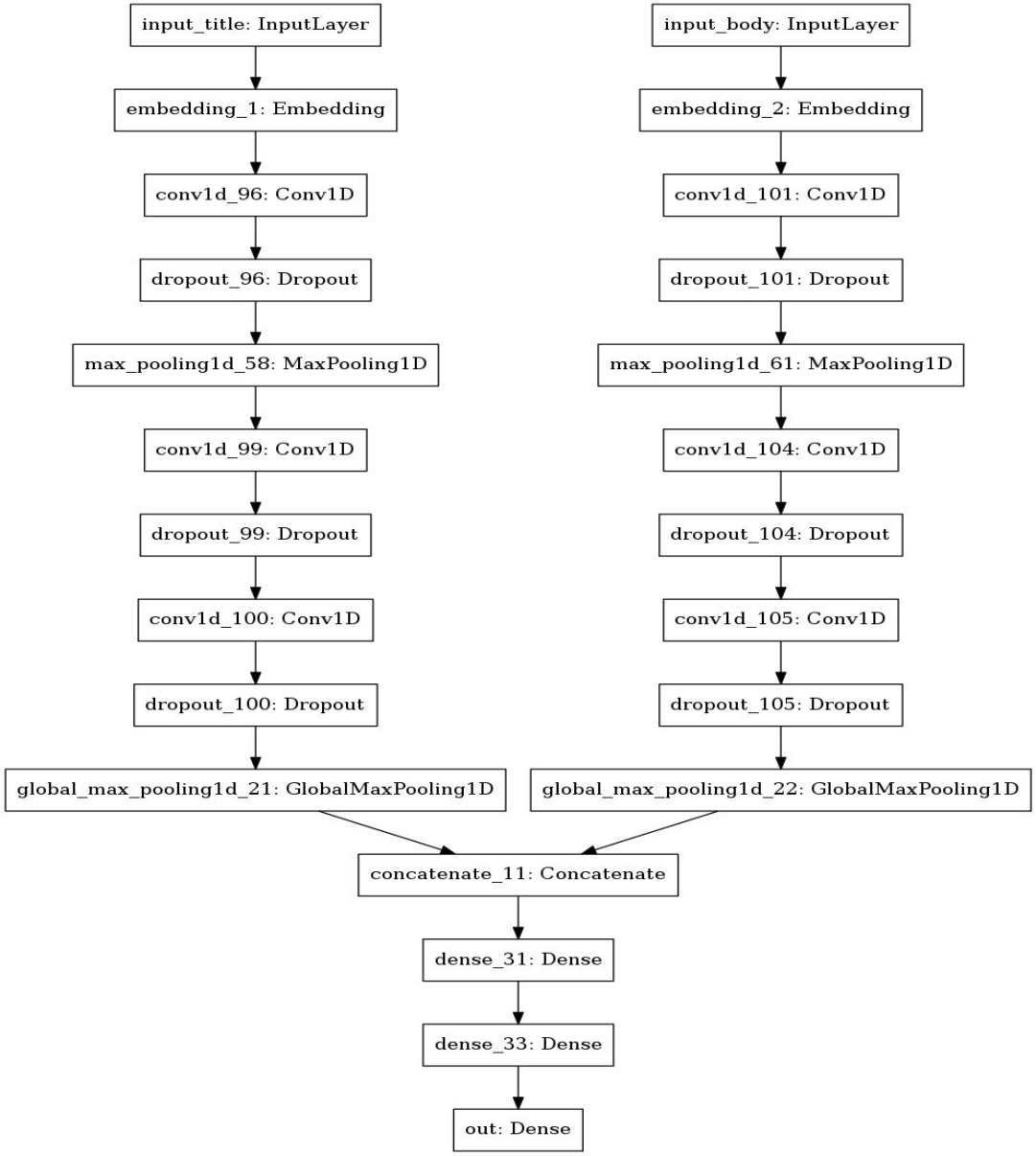
Parameters: 283,300

Results:

- Dev Accuracy - 84.1%
- Test Accuracy - 71.2%
- FNC Score - 4558

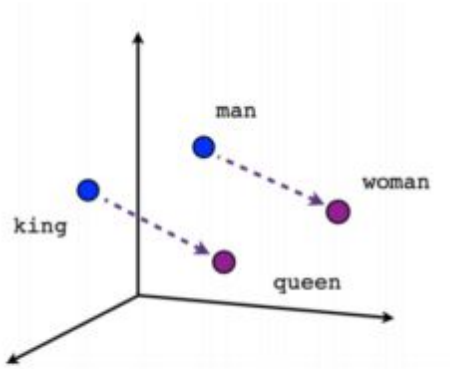
Future Development:

- Parameter tuning
  - Window Size
  - Pooling Size
- Softmax output layer split
  - Discuss Vs Unrelated
  - Agree Vs Disagree (Threshold)

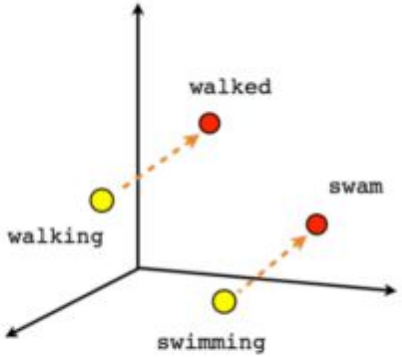


Bidirectional RNN encoder with Word Embeddings

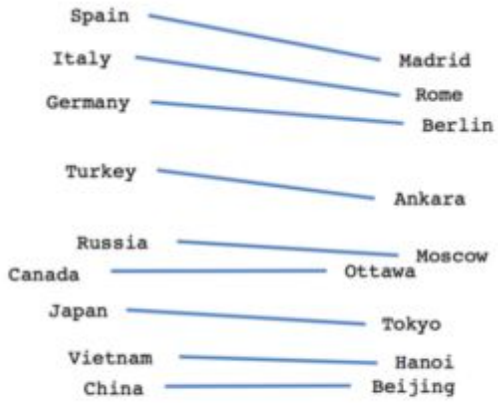
- Raw text is tokenized using NLTK
- Embedded using 50-dimensional word vectors using GloVe 6B
- Headline texts are padded with length 50 and body text padded with length 200



Male-Female



Verb tense

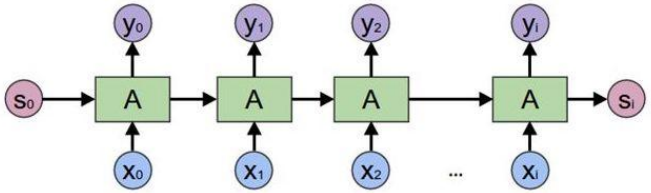


Country-Capital

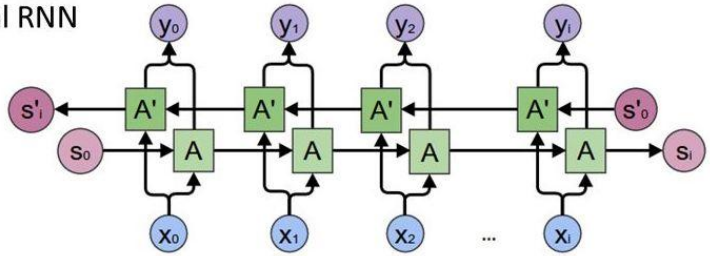
Bidirectional encoder with Word Embeddings

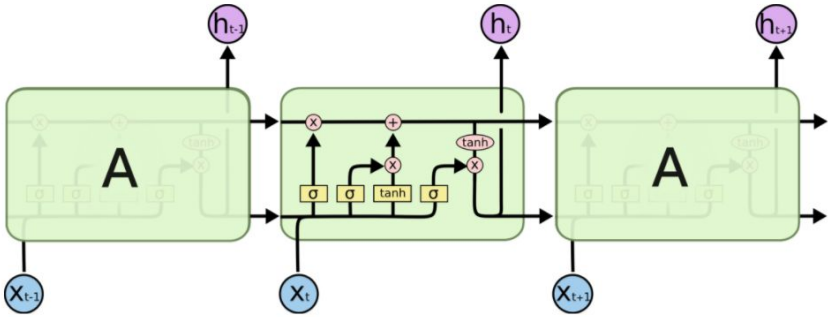
- The embeddings are then passed on to two separate LSTMs with 200 hidden units which then encode the two texts separately

RNN



Bi-directional RNN

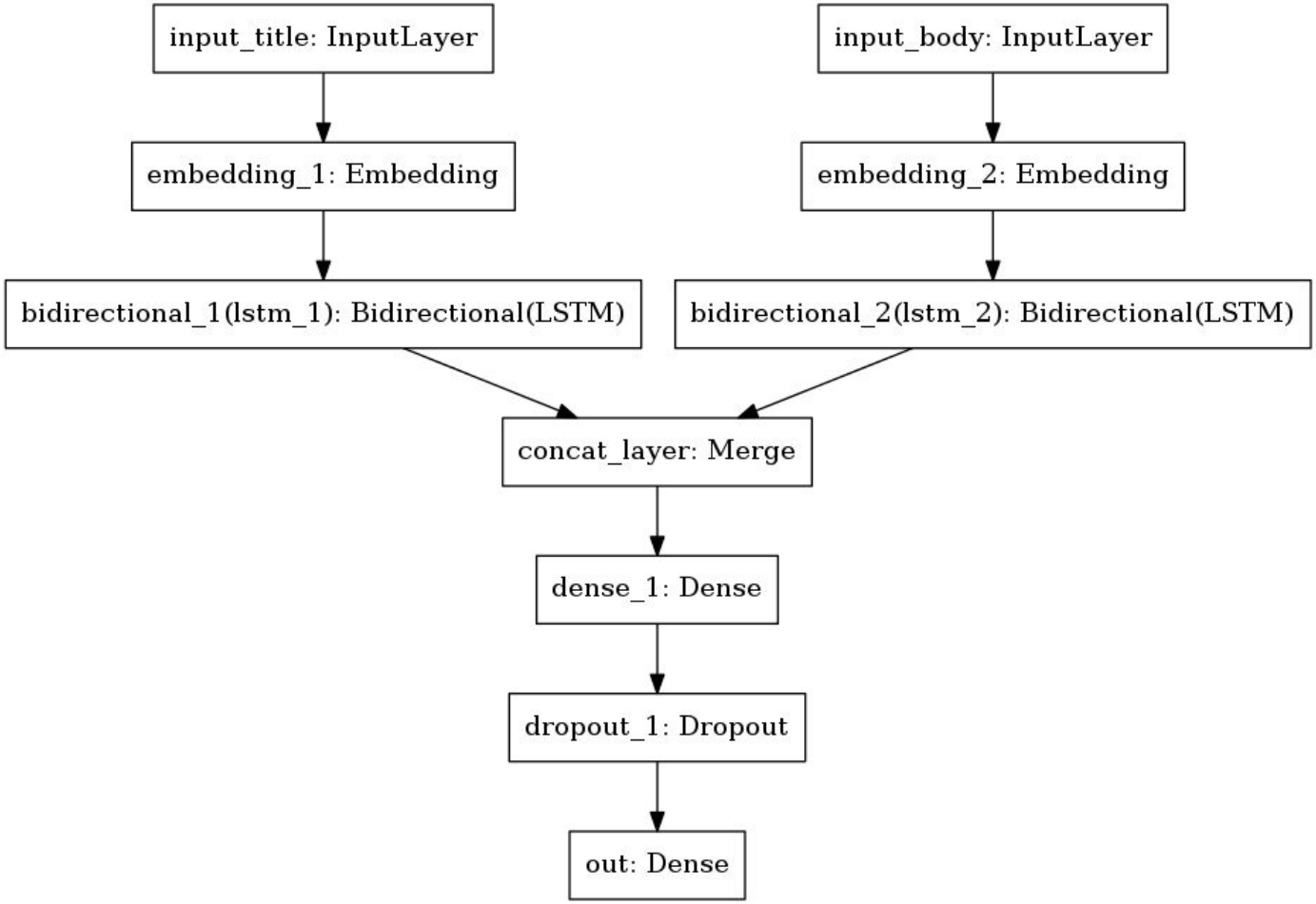




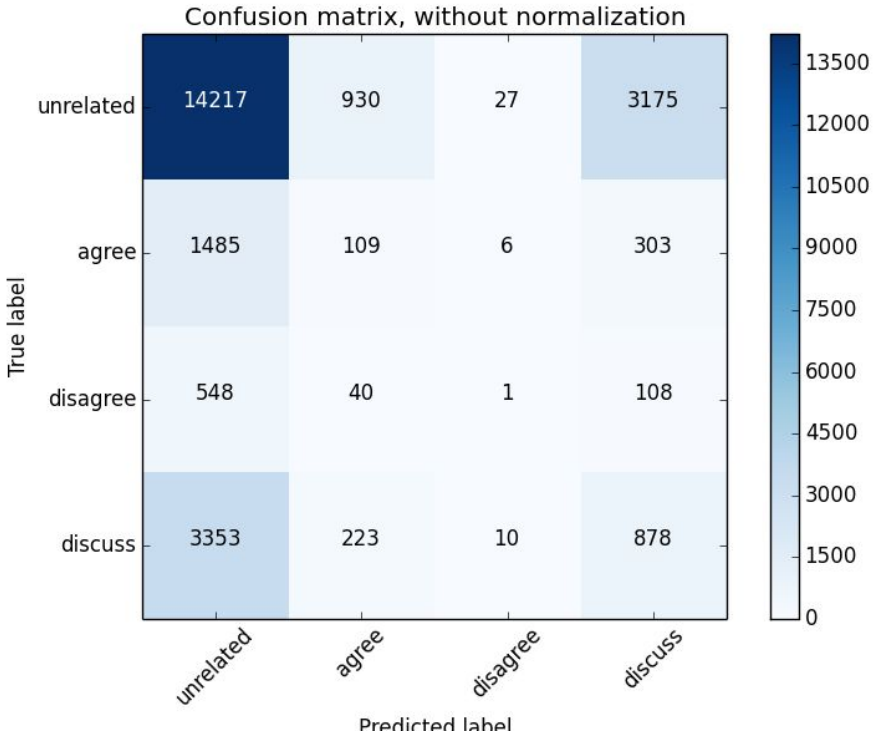
The repeating module in an LSTM contains four interacting layers.

- The encoded outputs are then concatenated and passed to the classifier which is a 1-hidden layer forward NN with 100 units in the hidden layer
- The final output layer of this forward NN is a 4-class softmax representing probability of 4 classes we are predicting here
- A dropout with keep probability 0.9 is applied only on the hidden layer of classifier

Model architecture



Results

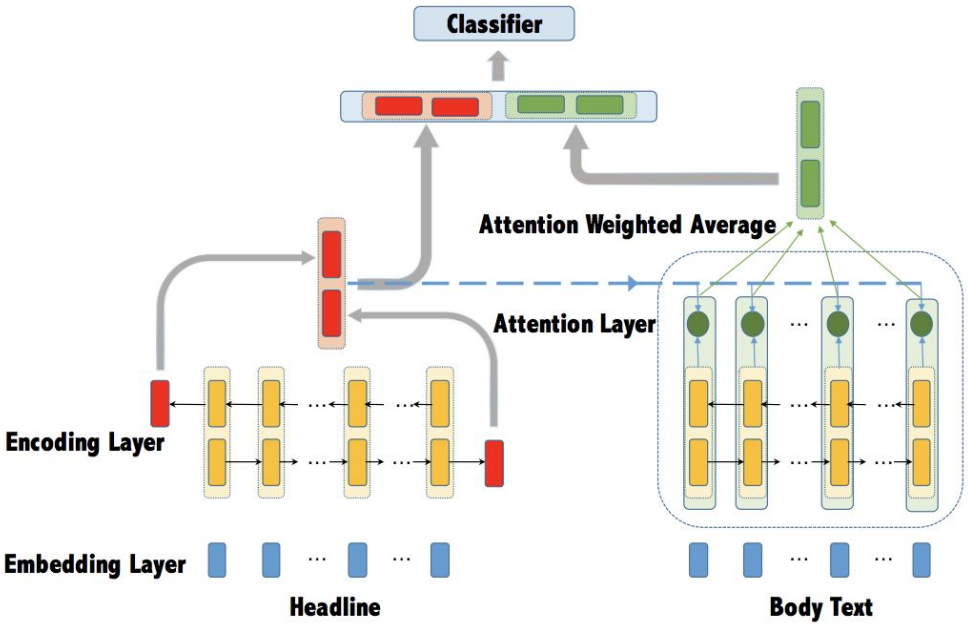


FNC score: 4714.75

Accuracy: 59.83

Future work

- Incorporating bias of the dataset into a customized loss function
- Introducing attention mechanism for finding the most relevant part in body text for a headline



Thank you!