

## News Headline Generation

### Overview:

Newspaper plays a significant role in our day to day life. In a news article, readers are attracted towards headline. Headline creation is very important while preparing news. Our goal is to implement text summarization by generating headline for a news body using recurrent neural networks.

### Data:

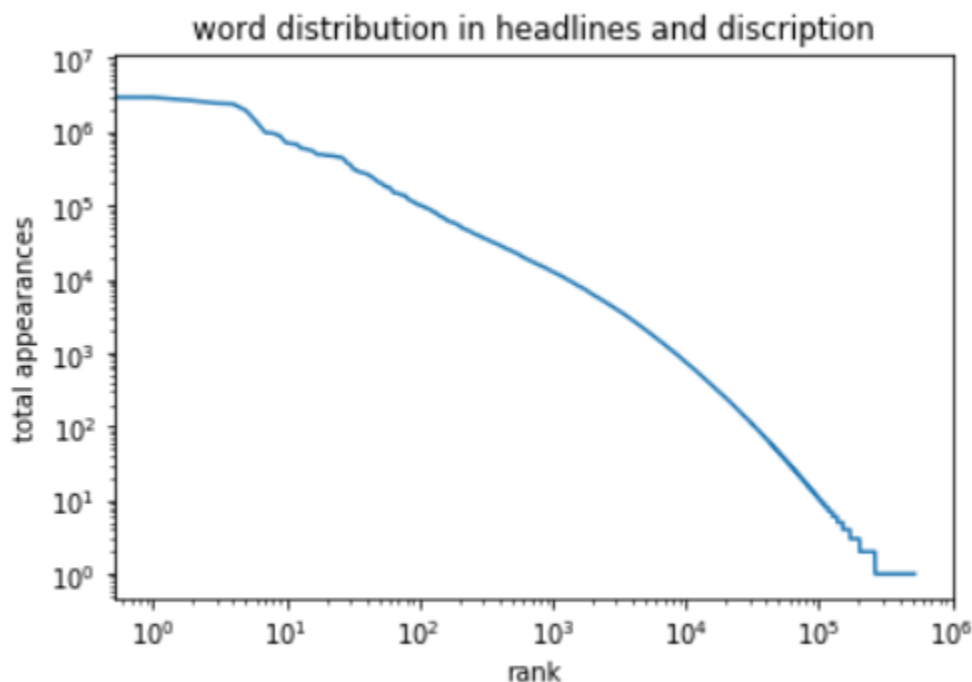
We have used “All the news dataset” from Kaggle. The data set consisted of approximately 1,50,000 articles. We extracted only title and content from the dataset.

We preprocessed our data from csv file by removing punctuations, stop words and converting the text lower case. Later, we tokenized the title, content and saved it as a pickle file.

We divided our data set into train (100000 articles), validation (29999 articles) and test validation (12567 articles) sets respectively.

We are using google news word2Vec file to map words in our data to vectors and to initialize weights in the model

Below graph depicts words distribution in headlines and articles



### Model Architecture:

Our model follows the common sequence-to-sequence learning framework with attention. It has three components: an encoder network, a decoder network, and an attention network as shown in the figure.

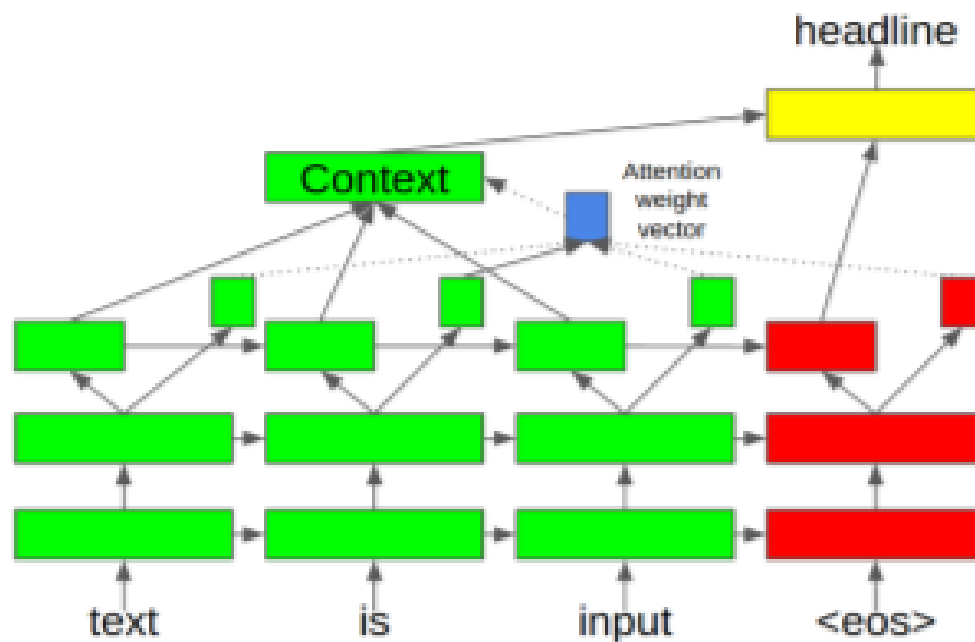
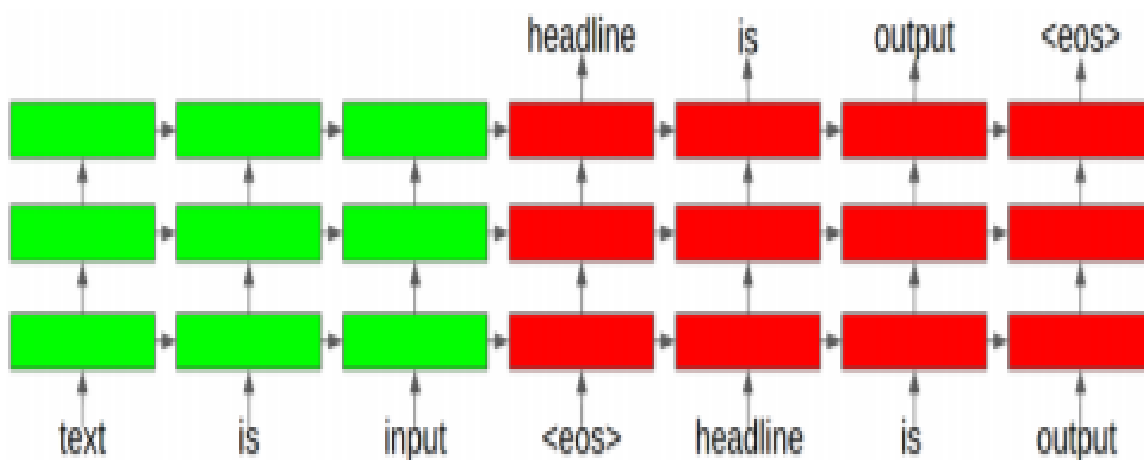


Figure 3: Simple attention



In our model news article and its corresponding headline is fed as an input to our encoder model.

The encoder model transforms the input into its vector representation. At the end of headline, we generate <unk> token to replace the word if it does not exist in our WordtoVec model, <eos> token to denote end of the headline and an empty token. We add paddings if the size of tokens is less than max size.

A decoder network then used this vector representation to output words until the end of sequence token was reached. The model stops making predictions after outputting the end-of-sentence token. The LSTM network inside our decoder model reads the input sentence in reverse, because doing so introduces many short-term dependencies in the data that make the optimization problem much easier. Decoder model then uses a SoftMax layer and attention mechanism to generate each word of a headline.

The final model was a combination of an embedding layer, 4 LSTM layers, simple context layer, time distributed dense layer and a softmax activation layer.

Beam search was used during the inference to generate the next word for our new headline

## Training:

Input sequences were reversed

Layer of top 40000 words out of 3 million words in google word2vec dictionary were used to represent the input words

<eos>, <unk>, <empty> tokens were appended at the beginning of 40000 words

SoftMax was used on the output layer

The model is fit for 20 epochs which took us 2 days to train

The batch-size of 128 sequences was used during training

The model was fit on Amazon AWS P2 (GPU) based instance

## Testing:

Load saved weights

Get headlines and descriptions from pickle file (or do real time using app)

Perform beam search to calculate most frequently occurring words

Convert list to string and display sentence

## Result:

newstitle	Some Republicans grill Ryan over House chaplain firing
newsbody	At least two Republican House members confronted Speaker Paul Ryan on Friday morning at their caucus meeting over the firing of the House chaplain, multiple Republicans coming out of the meeting Friday told CNN. One of the members who brought up concerns was New York Republican Rep. Pete King, who said afterward that Ryan's explanation for asking the Rev. Pat Conroy, a Jesuit priest, to resign was "unsatisfactory." Both King and Ryan are Catholic.
Generated Title	firing chaplain House over Ryan grill Republicans Some The Trump is the New York on Trump The the
Cosine Similarity	0.5773502691896258

## Challenges faced:

Web Scrapping/ preprocessing the data

Resource exhaustion – large word2vec, large batch size, attention mechanism, etc

Time taken to train model with bleu score

Integrating with web app/ training web app on aws

## Future work:

One can try using GRU based model instead of LSTM and train for more epochs.

Bi-directional RNN can also be considered since it provides more information to the model

## Conclusion:

We successfully implemented news headline generation app with encoder-decoder architecture along with attention mechanism. Our model generated headline that has cosine similarity of up to 0.5 while comparing it with original headline.

## References:

Research Paper: <https://arxiv.org/pdf/1512.01712.pdf>

Reference Code: <https://github.com/kabrapratik28/DeepNews>

RNN: <https://machinelearningmastery.com/how-does-attention-work-in-encoder-decoder-recurrent-neural-networks/>