Natural Language Processing

RNN Based article caption generator

Palash Parmar
626008848

Abhinav Ratna
227001723

Overview

- Problem Description
- Introduction
- Motivation
- Dataset and Preprocessing
- Methodology
- Evaluation and Analysis
- Results
- Conclusion
- Bottlenecks
- Future Work
- References

Problem Description

- Generate a condensed meaningful summary from an input article
- Article is a sequence of N words and the summary generated is a sequence of M words where M < N
- Generated summary should preserve information content and overall meaning
- Arduous and time consuming task for Human beings to summarize large documents of text

Introduction

- Text summarization is a task of generating a short summary consisting of a few sentences that capture salient ideas of a text

- Recurrent neural networks based designs very effective for various transductional tasks such as Machine translation, speech recognition etc

- Implemented model comprised of encoder decoder RNN with LSTM units and attention mechanism

// Motivation

- Millions of news articles being published in a day by multiple sources with catchy headlines
- People generally make judgments based on reading the headline and skip the story
- Imperative to have a headline which captures prime motive of the story and provides useful and meaningful content
- Easier task for machines but painstaking for human beings

Dataset and Preprocessing

- Trained model on “SIgnal Media News Articles Dataset”
- Dataset consists of one million news articles from multiple publishers with clear delineated headline and text
- Headline and text are lowercased and tokenized with punctuation removed from words
- Kept only the first paragraph of text

Dataset and Preprocessing

- End of Sequence (EOS) added to both headline and text

- Articles with missing headline or text or having headline tokens more than 25 and text tokens more than 50 are filtered out

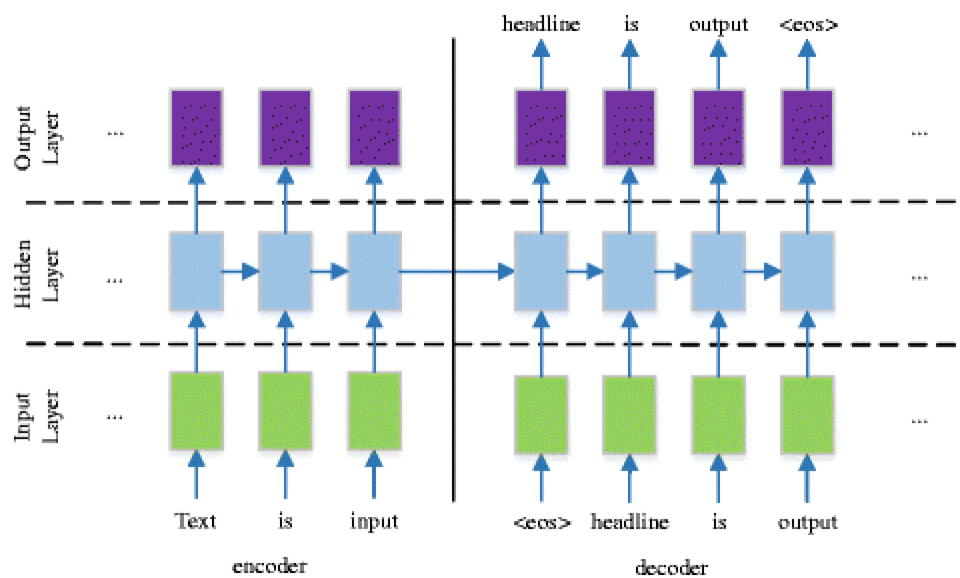
- All rare words are replaced with <unk> symbol keeping only the top 40000 most frequent occurring words

Methodology

- Architecture consists of two parts:
- Encoder and Decoder

- Both encoder and decoder are recurrent neural networks

- Used LSTM cell units
- for encoder and decoder
-
- Known as Sequence
- to Sequence Model



Encoding text

- Each word passes through embedding layer which transforms the word into distributed representation
- Embedded word vectors are then fed sequentially as an input to Encoder (LSTM Cell)
- Information flows from left to right and each word vector is learned according to not only current input but also all previous words
- When the sentence is completely read, encoder generates an output and a hidden state

// Decoding text

- Decoder (LSTM Cell) takes as input the hidden layers generated after feeding in the last word of the input text

- End-of-Sequence (EOS) symbol is fed in as input first using an embedded layer

- Decoder generates the text summary using a softmax layer and attention mechanism

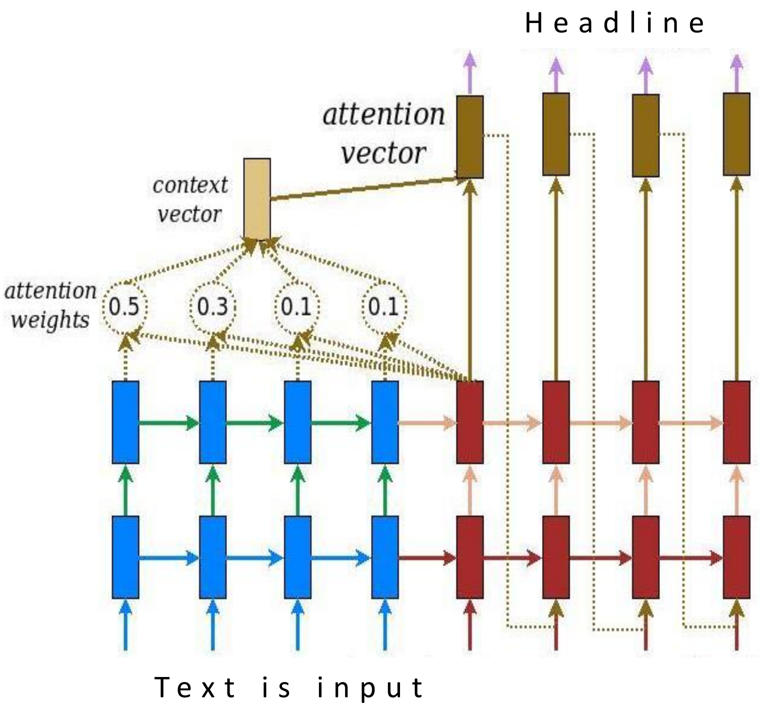
- Trained by teacher forcing (a mode that takes previous cell's output as current input)

Attention Mechanism

- Alleviates the need of encoding the full source sentence into a fixed-length vector
- Helps network remember certain aspects of the input better, including names and numbers
- Used when outputting each word in the decoder
- For each output word, it computes weight over each of the input word which determines how much attention should be paid to that input word

// Attention Mechanism

- Weights sum up to 1
- Used to compute a weighted average (context) of the last hidden layers generated after processing each of the input words
- Context is then input into the softmax layer along with the last hidden layer from the current step of decoding

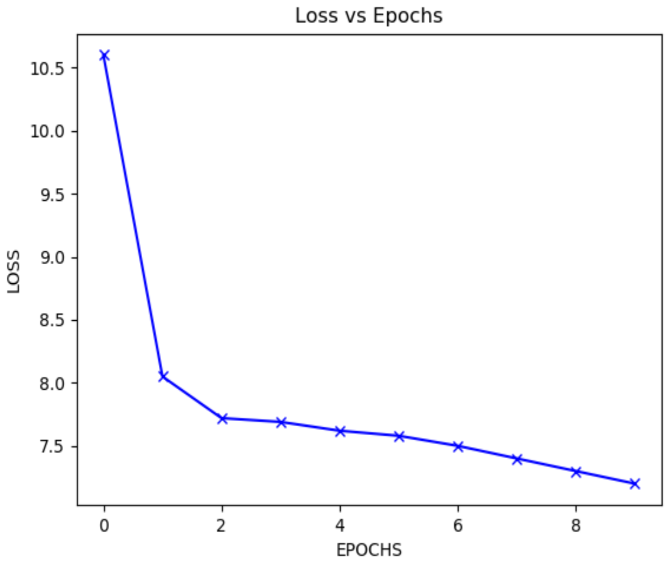


Evaluation and Analysis

- Measured the performance of model in two ways:
  - Training and test loss
  - BLEU (an algorithm for evaluating the quality of text) metrics over test examples

- Average BLEU Score for 400 test examples over 10 epochs -> 0.06

- Performance on test examples worse than training examples



// Results

Article	Actual Headline	Generated Headline
'We're too broke to go on the beat': Police chiefs warn ministers that cuts mean they can't do their job as they threaten to stop street patrols and say they won't be able to protect public from rioters or terrorists?	Police chiefs warn ministers Were too broke to go on the beat	Police chiefs say not able to protect public warn ministers
What have you been listening to this year ? If you want to find out using cold , hard evidence , then Spotify 's new Year in Music tool will tell you .	Spotify Will Make You Smarter for Your App	The 10 Most Popular Songs Of All Time
NHS patients to be given option of travelling to Calais for surgical procedures NHS patients in Kent are set to be offered the choice of travelling to Calais for surgical treatments, local health commissioners have confirmed.	NHS patients to be given option of travelling to Calais for surgical procedures	Patients travelling Calais

// Conclusion

- Model successfully generates concise summary for first 50 words of articles (taken as input)
- Valid and grammatically correct summary most of the times
- Attention mechanism in the model makes it easier to study the function of network
- Network learns to detect linguistic phenomena, such as verbs, objects and subjects of a verb, ends of noun phrases, names, prepositions, negations etc.

// Bottlenecks

- Dataset Preprocessing
  - Many training examples contain the headline which does not summarize the text at all or properly
  - Many examples have headline in the coded form such as "asn-rocr-2stld-writethru-sut"
- Model Training
  - Took 3 days to train the model on Nvidia Tesla K80 GPU for 200 iterations

// Future Work

- Using other type of attention mechanism called "Complex Attention"

- Implementation of Bidirectional RNN which is seen to be working better with attention model

- Use smaller dataset with better feature extraction to reduce the training time

- Use of GRU cell (faster than LSTM) to further reduce the model training time

// References

- Ian Goodfellow, Aaron Courville, and Yoshua Bengio. Deep learning. Book in preparation for MIT Press,2015.
- Bing L, Li P, Liao Y et al (2015) Abstractive multi-document summarization via phrase selection and merging[J]. arXiv preprint arXiv:1506.01597.
- Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. Scheduled sampling for sequence prediction with recurrent neural networks. CoRR , abs/1506.03099, 2015
- Cao Z, Li W, Li S et al (2016) Attsum: joint learning of focusing and summarization with neural attention[J]. arXiv preprint arXiv:1604.00125
- Andrej Karpathy and Fei-Fei Li. Deep visual-semantic alignments for generating image descriptions.CoRR, abs/1412.2306, 2014.

Questions?