CS564 - Text Mining
An Abstractive Text Summarization Using Bidirectional LSTM

Shakeel Ahmed
I181415

1 Problem Statement

Text summarization is a technique of producing a short, concise and fluent summary of a text while preserving key information and overall meaning of that text. There are two different approaches used for text summarization. The first is extractive summarization in which we point out important sentences from the original text and extract it. The second is Abstractive summarization, for which we are going to purpose methodology. Abstractive text summarization is the method of constructing summary sentences by merging facts from different source sentences and merging them into a shorter and concise representation while preserving all information of content and overall meaning. In abstractive text summarization, all the phrases and sentences are used to capture the meaning of the text document are entirely new and the sentences and phrases used in the original text documents are rarely used. For that reason, the abstractive methods for text summarization are more complex as compared to extractive text summarization. As the machine has to read over the text document and have to understand certain concepts and words to be important and preserve important meaning. Then the machine fetches some cohesive phrasing of the relevant concepts. Abstractive text summarization is similar to how we being human to summarize, as often we summarize by paraphrasing. In the course project, we are going to use a bidirectional LSTM based neural network approach for abstractive level summarization of text. For the project, we are going to use cnn/dailymail dataset available publically. We hope our purposed approach will outperform many previous approaches.

2 Motivation

Text summarization is a method of producing a short and fluent summary consisting of a few sentences that can capture silent ideas of a text. This is an important task in the domain of natural language understanding. a concise and good summary of a document can help humans comprehend the text document better in very less time. My motivation for this project is from my own experience. As a student, we often faced with a large number of scientific papers and research articles that are very important for us and also belong to our interests, but we can't read them due to the length of the document or short time. So by implementing this project, I want to be able to get the summaries of main ideas of the research papers, without significant loss of content. Text summarization is a widely used and implemented algorithm but I want to use a different approach that can be applied to scientific papers and articles as well.

3 Background

As the text summarization is a very interesting task. As being a student, instructor, and researcher, it is a very important thing to be used for text summarization. For a person to understand the problem one educated person should have familiar with the scientific writing and the pros of summarization and should have known the difference between abstractive and extractive text summarization. He also should have an idea of how much time is required to read a research article or a scientific paper. I hope being familiar with these things will be able to understand the importance of abstractive level text summarization.

4 Related Work

Text summarization is widly being studied now a days. kageback et al and [5] used neural networks to map vectors and then selected sencences based on theses vectors. [2] firstly applied neural networks in the abstructive text summarization, and got state of the art performace on DUC-2004, gigaword data set and two sentence level summarizing data set.This approach which was centered on attention mechanism was augmented with recent decoders [1], abstract meaning representations was studied (Takase et al.,2016)[4]. however the large scale datasets for summarization of longer text are rare. Nallapati et al.(2016) used the deep mind question answering data set (Hermann et al.2015) for summarization, resulting in CNN/Daily Mail dataset and output the first abstractive baseline. [3]used the pointer generater networks for text summarization.

5 Proposed Work

In this work, we are going to use bidirectional lstm for abstractive level text summarization. For that firstly we have to do world embedding. Word embedding is a popular representation of documents vocabulary capable of capturing the context of a word in a document. It is a combined name for a set of language modeling and feature learning method used in natural language processing. In word embedding the words and phrases from vocabulary are mapped to vectors of real numbers. There are many algorithms in world2vec library for the encoder input sequence. For purposed work, we are going to use skip-gram algorithm for the encoder input sequence. The skipgram for a given set of sentences or phrases (also called corpus), the model loops on the words of each and every sentence and phrases and either try to use the current word to predict its context or uses each of these contexts to predict the current work, due to which it is called skip-gram model. Actually, it uses each and every word of a large corpus and also use one by one word that are neighbors of it within a defined window and then feed a neural network that after training will predict the probability for each word to actually appear in the corpus. We use the skip-gram model for the encoding input sequence. For this, we will train a shallow neural network to predict the context words given a current word. And after training the hidden layer will be used as the embedding layer. we will pre-train the skip-gram model on articles and golden summary words.

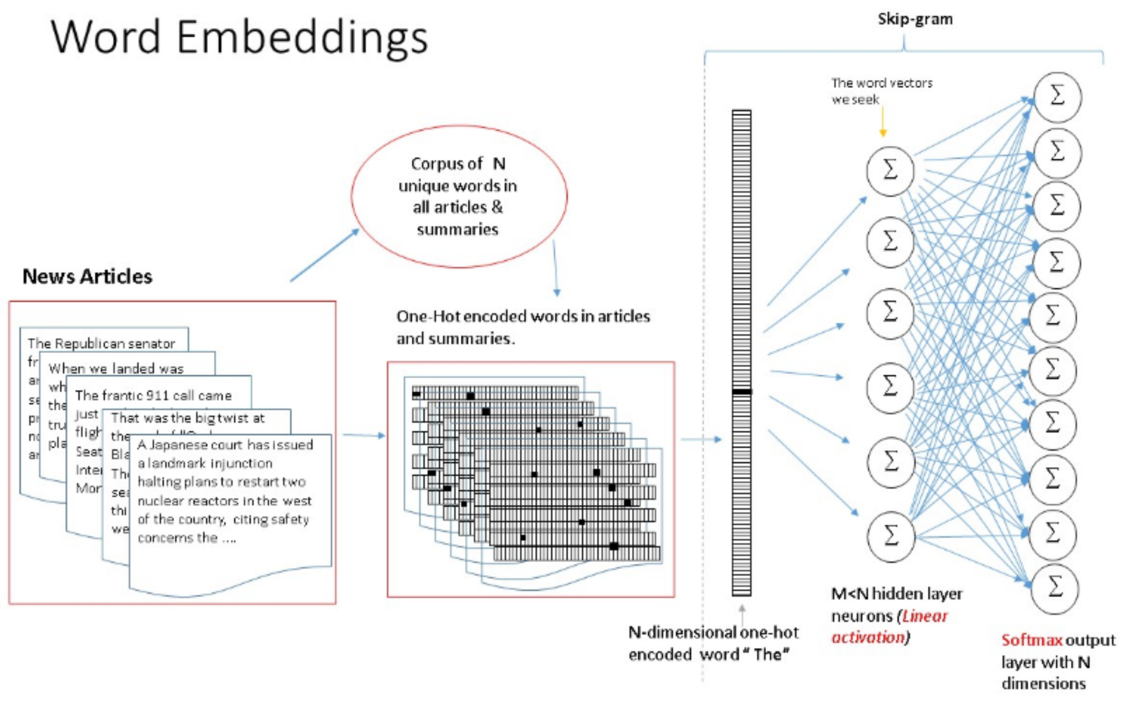
Word Embeddings
Skip-gram
The word vectors we seek
Corpus of N unique words in all articles & summaries
News Articles
The Republican senator
When we landed was
The frantic 911 call came
That was the big twist at
A Japanese court has issued a landmark injunction halting plans to restart two nuclear reactors in the west of the country, citing safety concerns the ....
One-Hot encoded words in articles and summaries.
N-dimensional one-hot encoded word "The"
M<N hidden layer neurons (Linear activation)
Softmax output layer with N dimensions

Figure 1: Word embedding and skipgram model

For a sequence to sequence prediction, we will use a bidirectional lstm encoder-decoder network. This approach is divided into two stages. one for reading the input sequence and then encoding it to a fixed-length vector and the second is to decode the vector into the predicted sequence. In the first phase, we will use the bidirectional encoder. This is presented in figure 2 here.

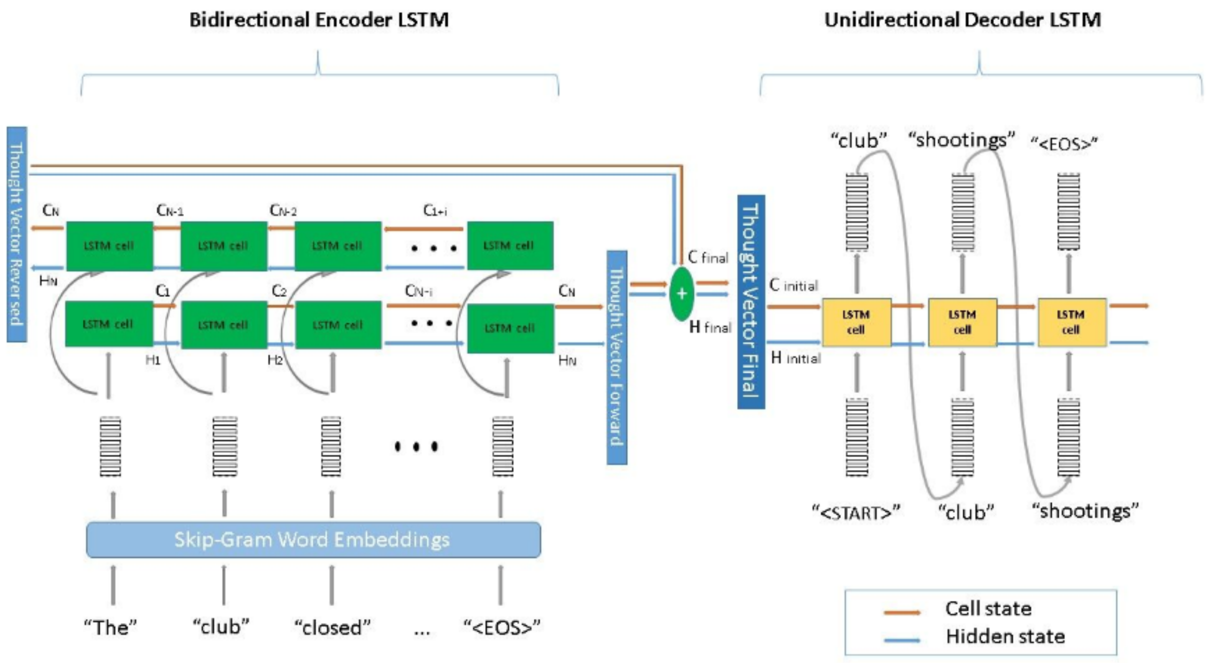

Figure 2: Encoder and Decoder LSTM

One attention layer between encoder and decoder over the source sequence's hidden state can be used if there will be the difference between the skip-gram embedding and one-hot vector sizes allowing multiplications with one-hot vector to get attention weights and vectors. attention layer can also be used in the output decoder sequence so that it can help to get decoder individual encoder state info. As a result, we are aiming to output the shortest summary of the document having all the important information. For this reason, if we cant outperform we will go to make our model more complex and will add more layers of network and will go to add another sequence to sequence model to produce comparing results.

6 Evaluation Methodology

ROUGHE matrix stands for recall oriented Understudy for Gisting Evaluation. This matrix is used for the evaluation of automatically summarization of texts and also for machine translation. It compares the result of automatically produced model summary and translation of a set of reference summaries (mostly human-produced). we are aiming to evaluate our model with ROUGHE matrix's f1 score. If we will be on time, we can also go with the Meteor matrix for evaluation. We will compare our results with our base paper and some recent papers to get the results of our model. ROUGHE matrix used an F1 score for its evaluation. It also has precision and recall criteria for text summarization and machine translation-related tasks.

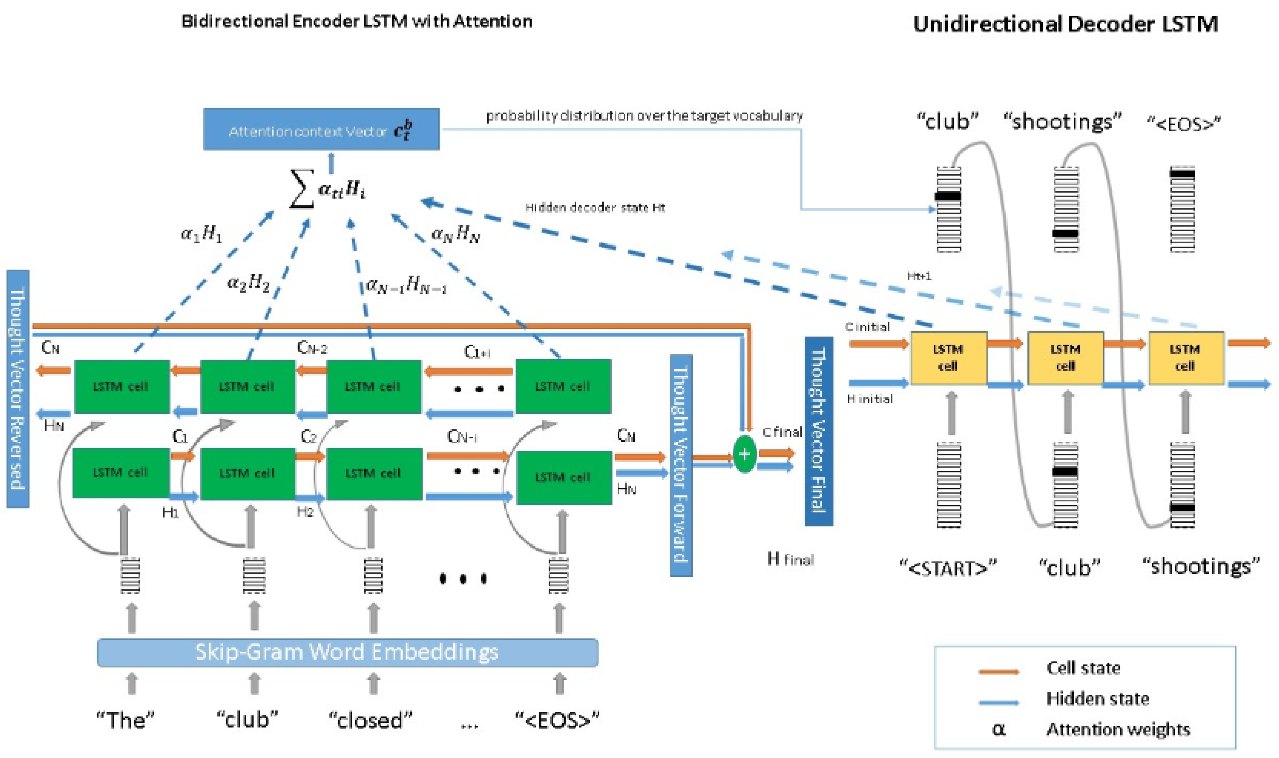

Figure 3: Encoder and Decoder LSTM with attention layer

7 Hypothesis

For the text summarization task we are going to develop a model based on bidirectional lstm. The purposed model is a bit complex and its evaluation is also challenging as its implementation. For the output, we are aiming to outperform our base paper results. For this time we cannot say how much we will be the result of our model, for time being if we are unable to produce the best result we will go to make our model more complex and will go to add more network layers to produce the best result as compare to our old papers.

8 Proposed Timeline

The tentative weekly timeline giving concrete milestones would be as follows. Modify the following timeline for your project.

- **October 6**: This proposal is submitted.
- **October 24**: Implementation of the model will be on the go.All installations will be done upto that time.
- **October 30**:we will aim to implement skip gram model. And dictionary corpus will be selected.
- **November 7**: We will be implementing the bidirectional lstm encoder and decoder, which will be our main challenge.
- **November 14**: we will implement our models roughe matrix or meteor matrix or may be both, depending on the time.
- **November 21**: We hope will will be completing our implementation if all will be on time.
- **November 27**: We will write our final report in this week that will be lengthy and so challenging.
- **December 5**: In the last week we will be ready for presenting our model and we will submit our final report.

References

[1] Sumit Chopra, Michael Auli, and Alexander M Rush. Abstractive sentence summarization with attentive recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 93–98, 2016.

[2] Alexander M Rush, Sumit Chopra, and Jason Weston. A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685*, 2015.

[3] Abigail See, Peter J Liu, and Christopher D Manning. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368*, 2017.

[4] Sho Takase, Jun Suzuki, Naoaki Okazaki, Tsutomu Hirao, and Masaaki Nagata. Neural headline generation on abstract meaning representation. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 1054–1059, 2016.

[5] Zhaopeng Tu, Zhengdong Lu, Yang Liu, Xiaohua Liu, and Hang Li. Modeling coverage for neural machine translation. *arXiv preprint arXiv:1601.04811*, 2016.