News Article Summarization Using RNNs

Ayush bhagta *IIT2019501* | Mohammad Monish *IIB2019033* | Sanket Kokude *IIB2019034* | Harshit Kumar *IIB2019035* | Viful Nirala *IIB2019036*

***Abstract*—The approach to text summarization has evolved from extractive to abstractive with the advancement in deep learning techniques. In this paper, we propose a Recurrent Neural Network based model that can be used in news article summarization. This idea is a natural improvement over the feed forward neural network based model. Due to inability of the feed forward neural network to process the output, the RNN does a nice job of re-feeding the output into the input and improve over many iterations. Further improvements proposed to this traditional RNN are inclusion of Long Short Term Memory RNN with encoder-decoder architecture with Attention. The limitation of this summarization technique is that it won't work for summarization in different languages. The future work might include more sophisticated tools in deep learning and large data collection to handle such drawbacks.**

Index Terms—Recurrent Neural Network; Summarization;

I. Introduction

Text Summarization is an important branch of NLP that has piqued curiosity in the academia as well as information and content-based industries. From NLP point of view, summarization should not be limited only to clustering of related words. Today's summarization techniques should also encompass the semantic values and be able to derive human-understandable meaning from it. Automatic text summarization has many use-cases, for example, many information present on the internet is redundant and might not be useful for the reader. We can use summarization to extract the meaningful information out of any information and save time. Other usage includes one-line news heading summarization, code summarization, structureddata summarization, etc. Extractive and abstractive are the two types of summarizations used for the above stated purposes. A subset of words are selected retaining the important points in the article during extractive summarization, whereas abstractive summarization is based upon the understanding the semantic of the input. We intend to use an abstractive approach for the project

In today's busy world, people complain about not having time to read a 7-8 page long newspaper and headlines don't provide enough information on the topic which people are interested in. Our aim is to summarize each article under a headline providing the most relevant/meaningful information present in the article. This implementation can be used for many other purposes such as: one-line news heading summarization, code summarization and structured data summarization.Taking these things under consideration we have used Recurrent Neural Networks (RNN). RNNs are mainly used for solving temporal or ordinal problems such as NLP and language translation. We further explored the encoder-decoder model with LSTM cells and Attention.

Our dataset is 'News Summary Dataset from Kaggle', which contains approx 100,000 news articles and their titles in the form of a csv file.

II. Literature Survey

Automatic Text Summarization, or the reduction of a text to its essential information, is a difficult topic that, despite recent advances, continues to offer several problems to the scientific community. Given the exponential expansion of textual material online and the necessity to quickly analyse the contents of text collections, it is also a relevant application in today's information society. Significant amount of past work done in summarization has been extractive methods, which focuses on identifying important sentences in the news to present them in a summary. However, humans on the other hand summarize the original news in their own words and seldom reproduce the original sentence from the news which semantically summarizes the text or rephrase the text in the corpus instead of quoting it verbatim. This method of summarization is abstractive in nature and is relatively new research area. With the incremental use of deep learning in many NLP tasks, abstractive methods of summarization have become a popular field. In 2015, Cho et al described state-of-art performance of attention-based encoder-decoder networks, for which output possesses the same structure of input [1]. In the same year, Rush et al developed a neural attention feedforward model for sentence-level summarization task which performed well on the DUC-2004 competition [2]. In 2016, Chopra et al designed attentive recurrent neural networks to improve the results on the same task [3]. The models we implemented are inspired by these three papers.

III. Problem Statement and Objective

Problem Statement: Given a news article, generate onesentence summarization that mimics the style of news titles.

Objective: Generation of one-line summary of news articles which captures the true meaning of the news article and that mimics the news titles. Here, we attempt to use basic LSTM based encoder-decoder network for news summarization and build complex models on the top of this basic encoderdecoder architecture and see how they perform against various evaluation metrics.

IV. Proposed Methodology

Our proposed method include incremental addition of complex models in the basic RNN architecture. The main

idea is that we will build a LSTM based encoder-decoder Network that will act as our base model. After that we'll add Attention to the architecture:-

Pre-Processing :

- Convert everything to lowercase
- Remove HTML tags
- Contraction mapping
- Remove ('s)
- Remove any text inside the parenthesis ( )
- Eliminate punctuations and special characters
- Remove stopwords
- Remove short words

Analyzing and Preparing the Data : We will analyze the length of the reviews and the summary to get an overall idea about the distribution of length of the text to fix the maximum length of the sequence
We observe that 98.8% of the summaries have length below 12. So, we can fix maximum length of summary to 12. Let us fix the maximum length of article to 50.

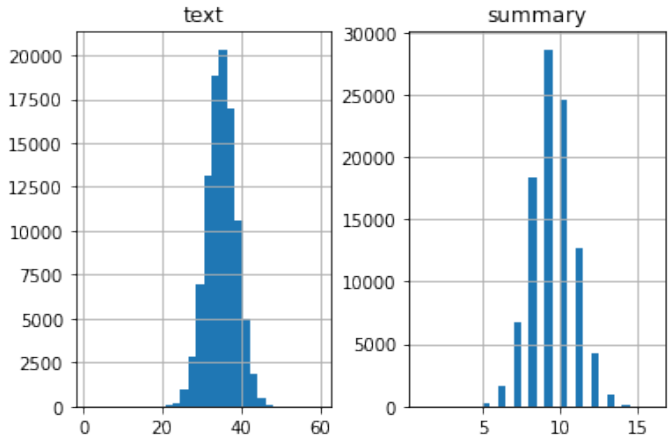

Fig. 1. Analyzing the Data.

1) Long Short Term Memory (LSTMs) : These are special kind of RNNs that can retain and map long term dependencies. These models are powerful because they have the ability to append new information and delete existing information using gates. The first step in the model is whether to retain the past information or remove it. The sigmoid function takes input previous information and current information and decides on to delete if value is 0 and keep it if value is 1. We will update the old information in the next step through a tanh activation function. This gets multiplied by the output from the sigmoid function and generate the required outputs.

2) Encoder-Decoder Network : An encoder-decoder framework is a general framework based on neural networks

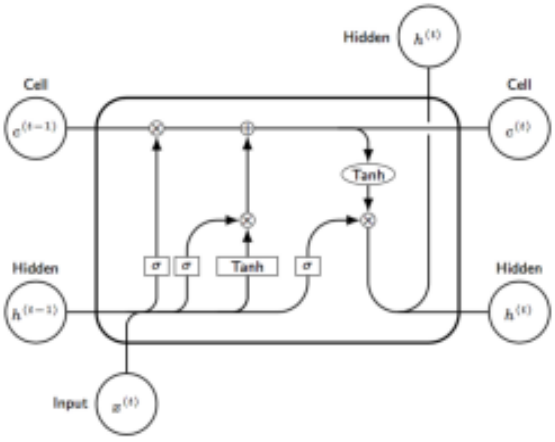

Fig. 2. Long Short Term Memory Architecture.

which consists of encoder that encodes the variable length input into fixed-length vector representations and decoder that decodes the fixed length vector representations into variable length output sequence. This encoder-decoder architecture can learn conditional probability distribution over a variable length sequence conditioned on yet another variable length sequence.

Encoder reads input sequence x sequentially and after reading the whole input sequence the hidden state of encoder unit generates fixed size vector representation c, which encapsulates the summary of whole input sequence. Similarly, decoder model generates output sequence y sequentially based on previous output y, hidden state h and vector representation c which was generated by encoder.

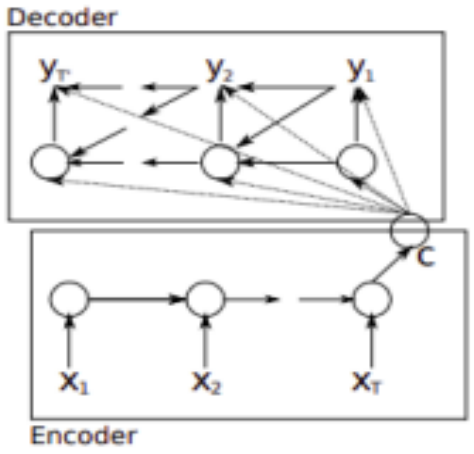

Fig. 3. Architecture of Proposed LSTM based Encoder-Decoder.

3) Attention Based Encoder-Decoder Model: Encoder produces an encoded vector which encapsulates all the information of the input sequence. This creates a information bottleneck here because we are trying to compress all the information of input sequence into this vector. Attention is used

to provide more focus to important parts of input sequence while generating the output sequence at each timestep. Fig. 3. Attention based Encoder-Decoder Model.

Instead of encoding the input sequence into single fixed vector, Attention allows for creation of context vector at each timestep based on decoder's current hidden state and encoder's hidden state. This context vector along with previous hidden state of decoder and previous decoder output is used to predict the current decoder's output. The context vector provides the attention distribution based on decoder's current hidden state and lets the decoder know where to attend to produce the target token.

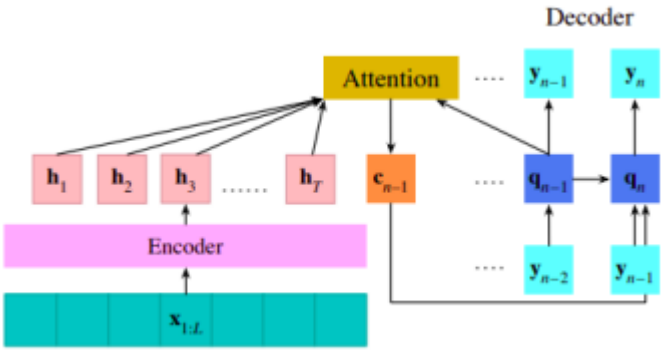

Fig. 4. Attention based Encoder-Decoder Model

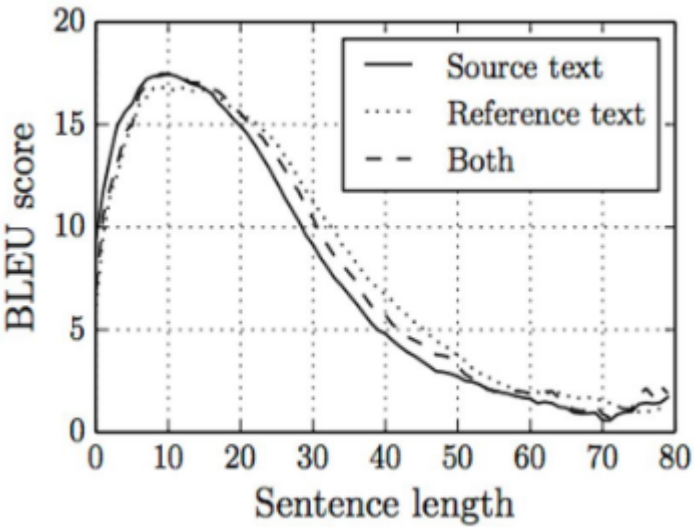

Fig. 5. BLEU score vs sentence length

Model Training

- We are using early stopping to monitor the validation loss metric. Our model will stop training if the validation loss starts increasing.
- We'll train the model on a batch size of 128 and validate it on the holdout set (which is 10

V. EVALUATION METRICS

1) BLEU : Bilingual Evaluation Understudy Score is a language independent and widely adopted metric for evaluating a generated sentence to a reference sentence. BLEU is similar to unigram precision, the only modification in BLEU is that for word in the generates text, BLEU takes its maximum count in reference text. In pracice, using unigrams/separate words isn't optimal so BLEU instead computes precision using n-grams of the word.

$$P = \frac{\sum_{n-gram} count_{clip}(n-gram)}{\sum_{n-gram} count(n-gram)}$$

2) ROUGE : In Natural Language Processing ROUGE is a set of metrics used for evaluating generated/auto summarized text to reference text. Recall-Oriented Understudy for Gisting Evaluation or in short ROUGE has five major evaluation metrics - ROUGE-N, ROUGE-L, ROUGE-W, ROUGE-S and ROUGE-SU. We have used ROUGE-1 and ROUGE-L for evaluating the summarized text which is generated by our model. During ROUGE-1 evaluation match rate of unigrams between generated text and reference summary is calculated. In ROUGE-L metric LCS (Longest Common Subsequence) between the two texts are taken for evaluation. In both the metrics precision and recall values are calculated and finally a score known as F1-score is calculated to measure the performance of the model.

$$F1score = 2 * \frac{precision * recall}{precision + recall}$$

VI. RESULTS

We train our LSTM based encoder-decoder network with attention for 30 epochs on 'News Summary dataset Kaggle'. We used early stopping to monitor the validation loss metric, so our model stopped training after 12 epochs when our validation loss started to increase. We have used considerably smaller model to reduce the training time on colab. We have not used any pre-trained word embeddings instead we have embedding layer in the model which is training with the model.

Review: two german users exploited twitter bug writing character long tweet instead usual characters users able post beyond word limit long message url twitter temporarily suspended users accounts confirmed exploit fixed
Original summary: twitter bug lets two users post character tweet
Predicted summary: twitter bug lets users read character in tweet

Review: dinosaur bird capable flying despite skeletal differences modern birds breaks according study journal nature birds evolved jurassic era small feathered dinosaurs represent dinosaurs survived mass extinction event million years ago noted researchers studied million year old fossil using lasers

Original summary: dinosaur bird could fly but in bursts study
Predicted summary: dinosaur bird could fly over years study

Review: launching swachhta hi sewa campaign union home minister rajnath singh friday said india expected open defecation free october providing toilet facility home ensure safety dignity women swachhta hi sewa campaign help improve nutrition productivity children rajnath added
Original summary: india will be open defecation free by oct rajnath
Predicted summary: india will open defecation free by oct rajnath

Review: us defence department decided seize artwork created guantanamo bay detainees longer release artwork public prisoners also longer allowed give red cross send families attorneys detainees claimed artwork destroyed
Original summary: us defence department to seize bay art
Predicted summary: us defence department to seize bay art

Review: brazilian football club chapecoense santa state championship sunday five months players died plane crash colombia november club three surviving players signed new footballers season began january club also find replacements scouts medics staff
Original summary: win title months after air crash killed team
Predicted summary: brazil team wins months after plane crash

Review: women iran shared videos dancing online show support arrested teenager hojabri iranian authorities arrested hojabri posted videos dancing instagram without wearing mandatory headscarf iranian state media released video hojabri acknowledged breaking moral norms
Original summary: iranian women dance to show support for arrested teenager
Predicted summary: iranian women dance to support teenager arrested

VII. CONCLUSION

In this project, we aim to apply various approaches for abstractive summarization of news article, generating onesentence summarization that mimics the style of news titles. We start from our baseline model which is basic LSTM based encoder-decoder architecture and then we add Attention to that encoder-decoder architecture and compare their performances with respect to actual reference. We believe abstractive text summarization is more human way of summarization and we will continue this approach in future. As part of our future work, we plan to focus our efforts on more state of the art

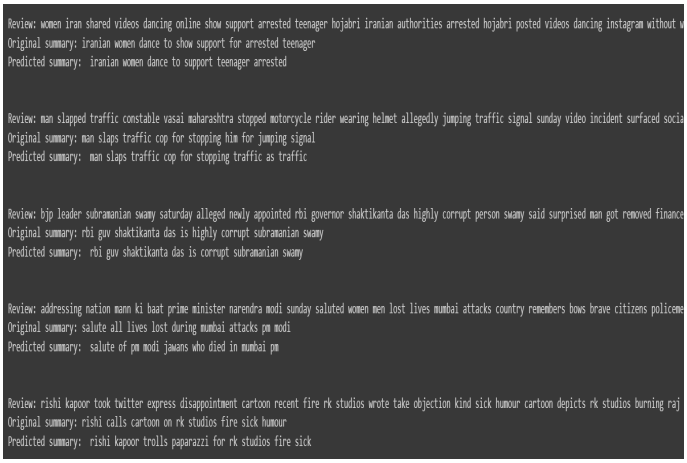

Fig. 6. Final predicted summary.

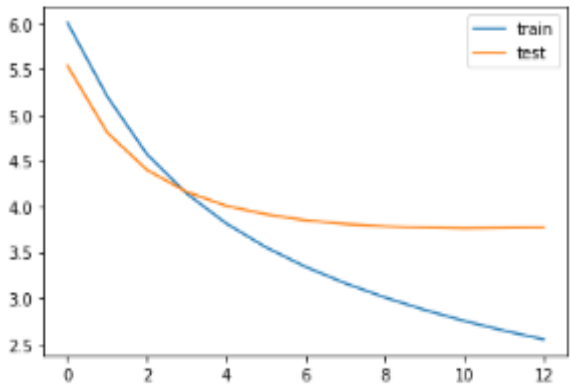

Fig. 7. Architecture of Proposed LSTM based Encoder-Decoder

models such as Transformer based networks and build more robust models for summaries consisting of multiple sentences.

REFERENCES

[1] Cho, K., Courville, A. Bengio. Y. (2014) Learning Phrase Representations using RNN EncoderDecoder for Statistical Machine Translation. 2014 Conference on Empirical Methods on Natural Language Processing (EMNLP).

[2] Rush, A.M., Chopra, S. Weston, J. (2015) A Neural Attention Model for Abstractive Sentence Summarization. 2015 Conference on Empirical Methods on Natural Language Processing (EMNLP).

[3] Chopra, S., Auli M. Rush, A.M. (2016) Abstractive Sentence Summarization with Attentive Recurrent Neural Networks. 2016 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology.

[4] Y. Wang, X. Fan, I.-F. Chen, Y. Liu, T. Chen, en B. Hoffmeister, “End-to-end Anchored Speech Recognition”, arXiv [cs.CL]. 2019.

[5] Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. In Text summariza- tion branches out: Proceedings of the ACL-04 workshop, volume 8, 2004.