

# RNN ENFORCED COMPUTERIZED SYNOPSIS BUILDER

Sweta Gupta

Dept of Information Technology

V.C.E.T

Vasai, India,  
[shwetagupta22031607@gmail.com](mailto:shwetagupta22031607@gmail.com)  
[m](mailto:shwetagupta22031607@gmail.com)

Isheet Shetty

Dept of Information Technology

V.C.E.T

Vasai, India,  
[isheetshetty@gmail.com](mailto:isheetshetty@gmail.com)

Yash Jobalia

Dept of Information Technology

V.C.E.T

Vasai, India,  
[jobaliayash@gmail.com](mailto:jobaliayash@gmail.com)

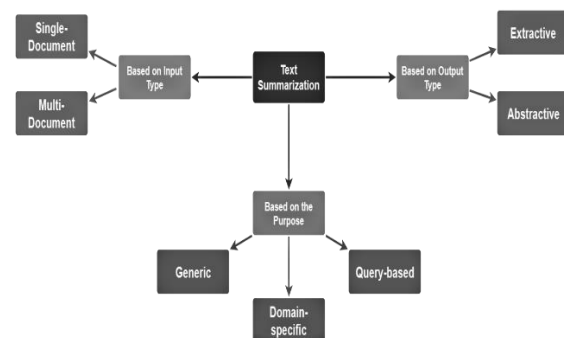
**Abstract**—Today, our world is parachuted by the gathering and dissemination of huge amounts of data. With such a big amount of data circulating in the digital space, there is need to develop machine learning algorithms that can automatically shorten longer texts and deliver accurate summaries that can fluently pass the intended messages. Hence, text summarization is the task of producing a concise and fluent summary without any human help while preserving the meaning of the original text document. So, to save time and easy to understand the meaning of the text summary is important. An automated summary is an effective way to solve this problem.

**Keywords**— Text Processing, Deep Learning, Bi-directional RNN, Encoding, Decoding.

## 1. INTRODUCTION

When you open news sites, do you just start reading every news article? Probably not. We typically glance the short news summary and then read more details if interested. Short, informative summaries of the news is now everywhere like magazines, news aggregator apps, research sites, etc. Well, it is possible to create the summaries automatically as the news comes in from various sources around the world. The method of extracting these summaries from the original huge text without losing vital information is called as Text Summarization. Text summarization can be classified on the basis of different criteria. Synopsis based on construction, extractive summaries pick important sentences from the document based on certain conditions and display it to the user as they are, abstractive summaries gives reconstructed summary which is not exactly the same as the original document. On the number of sources for the summary, single document summary which is produced from single document, multi-document summary which is obtained from multiple documents. Trigger based abstracts can be of two types, generic based summary, present the summary in concise manner as the main topic of the data and query based summarization gives synopsis as an answer to the

query given by the user. Also based on the important details of the synopsis it can be classified as indicative which gives information to the user whether the document should be read or not, and the informative abstract provide all the relevant information to represent the original document. The most challenging problem of auto summarization is to provide information that is relevant to user's topic of interest.



## II. MOTIVATION

Now days more and more information is available through internet and other sources. To handle these data more efficiently we need a tool for extracting proper set of sentences from the given documents. Summarization of text is essential to get the important information while dealing with large collection of documents. With the advent of World Wide Web information has become intrinsic part of our life. To remember the details of every information is not possible for human mind. Therefore summarization of text documents plays a very important role in information gathering. In this study we are using deep learning Algorithm for the summarization task. In today's world time is precious and equivalent to money, no one has the time to read long reports. Hence summarizer helps to save that precious time by creating short

summaries of long texts and chapters .

### III. REVIEW OF LITERATURE

This section describes about the methods that have been used for text summarization. In Natural language processing, text summarization is one of the fields. Text summarization is divided into 2 types: 1) Extractive summarization 2) Abstractive summarization. An Extractive text summarization is, choosing necessary sentence from the text. This important paragraph can be selected by using linguistic and statistical features of paragraphs. Abstractive summarization understands the main concept and meaning of the given document or text. It finds the new concept from the document by using linguistic method by interpreting the text.

In earlier researches, summarization was done on scientific documents based on the proposed features like sentence ranking. J.N.Madhuri and Ganesh Kumar.R have done their research on extractive summarization by extracting high weighted frequency sentences .In this paper we see that after completing the pre-processing step they calculate the frequency of each keyword like how frequently that keyword has occurred, from that maximum frequency of the keyword is taken. Then weighted frequency of the word is calculated by dividing frequency of the keywords by maximum frequency of the keywords. In this step, they calculate the sum of weighted frequencies. Finally, the summarizer extracted the high weighted frequency sentences,

Similarly, Aakanksha Sharaff, Amit Siddharth Khaire and Dimple Sharma analysed fuzzy based logic for extractive text summarization in which first the dataset is pre-processed which is used to calculate frequency and position. This is done by using hash map and formulae after which the weight of sentences is calculated which is done by using frequency and position of words formula. The last step is fuzzy analysis where the calculated values are used to find the mean for that sentence and are fed to the triangular membership function which gives values between 0 and 1 to each sentence and the fuzzy rule is applied. These values are arranged in descending order and they are picked according to the percentage of the original text needed. Finally, the ROUGE score is given to each and every summary to be compared efficiently.

In the next paper, Siya Sadashiv Naik, Manisha Naik Gaonkar [3] research about extractive text summarization by feature based sentence extraction using rule based. The main focus of this paper is to summarize a single document and create its

extractive summary. After pre-processing step, each sentence of document is represented as attribute vector of features. Seven features are calculated for each sentence and each feature is given a value from 0 to 1 after normalization. Features considered are Sentence Position, Title Feature, Numerical Value, Keyword Weight, Proper Noun, Sentence-To-Sentence Similarity and Sentence Length. All the sentences are sorted in ascending order based on their scores. Finally, the extractive summary of the document will be generated and displayed.

In the following paper, Kaiz Merchant and Yash Pande [4] use latent semantic analysis approach for creating short summaries on basis of similar words. They use 2 approaches depending on the type of case if it is a criminal case, they used single document untrained approach and for civil case they used multi-document trained approach. They first pre-processed the data which is cleaning it, lemmatisation and removing stop words then they pass it through the model and depending on the type of case (civil case or a criminal case) it is decided which process is used. Then using the appropriate model, they generate a summary and finally add sentence selection, in which the final line is always added because it is the judgement passed and hence the final summary output is generated.

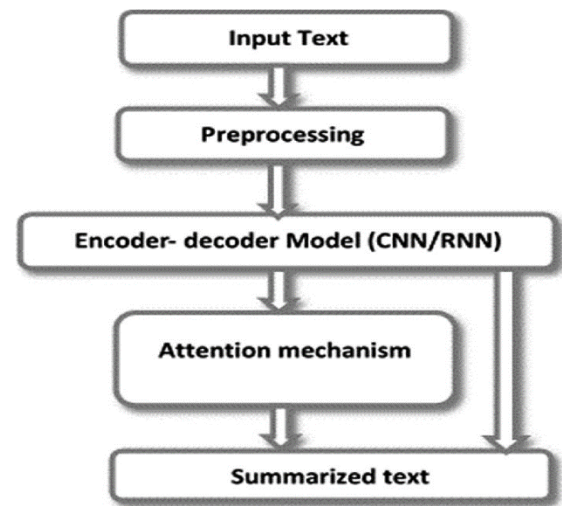
Abu Kaisar Mohammad Masum [5] described an end-to-end approach to sequence-to-sequence learning used a Multilayer LSTM. The neural network contains encoder and decoder. Encoder used a fixed length of text using as input and Decoder represents the output used tensor flow CPU version. Their output was positive gave short summary for few words.

### IV. IMPLEMENTATION

In this section, we represent our methodology for making an abstractive text summarizer for English-to-English text document. Previously there are many successful works were completed for making a text summarizer in English text. But we have tried to produce a better text summarizer using our own approach and we were able to create a better abstractive text summarizer for English text documents. The objective of the project is to understand the concepts of natural language processing and creating a tool for text summarization with machine learning containing only the main points described in the document. We first record the conversation or the meeting from a voice recorder on a phone or a mic then we use STT (speech to text ) algorithms which recognise speech and convert them to text as our first input this

converted text file will be then encoded and decoded by algorithms by either extractive summarisation or abstractive summarisation in order to convert the lengthy conversation or meetings into small summaries of the entire topic.

Hence we justify that abstractive text summarization creates a summary of a text document using its intrinsic and chooses the key content of the text document using potential vocabulary. A sequence of aim words as input text in a source text document and predicted aim words of a sequence is called a summary of a text document. For a short text, summarizer has established a self-encoder, decoder RNN attention model on machine translation to text summarization describe an end-to-end approach to sequence to sequence learning used a Multilayer LSTM. The neural network contains encoder and decoder. Encoder used a fixed length of text using as input and Decoder represents the output. The sequence to sequence learning has done improved neural machine translation. There are two ways of attention model one global and another is local. Global approach accepts all of the source text words and local accept the only subset of source text at a time. Both approaches are effective for machine translation. Neural network response to producer short text conversion. When we read a long text its kills our time. Some time it is not easy to understand the meaning of this text. So, to save time and easy to understand the meaning of the text summary is important. An automated summary is an effective way to solve this problem. The main motivation of our research is to generate automate abstractive text summarizer to generate summary easily by machine. Since a build in the model used in our experiment but if increase the efficacy of the model processing text and model input and output is important. This paper we were discussing our contribution in coding site to increase the accuracy and reduce the loss in train time. We introduce an easy to pre-process and apply lemmatization to reduce the inflectional form on the word in summary.



#### IV. REFERENCES

1. J.N.Madhuri , Ganesh Kumar.R “Extractive Text Summarization Using Sentence Ranking”, Institute of Electrical and Electronics Engineers (IEEE), 2019
2. Aakanksha Sharaff, Amit Siddharth Khair, Dimple Sharma,” Analysing fuzzy based approach for extractive text summarization”, International Conference on Intelligent Computing and Control Systems (ICICCS 2019)
3. Siya Sadashiv Naik, Manisha Naik Gaonkar,” Extractive Text Summarization by Feature based sentence extraction using rule based”, IEEE International Conference On Recent Trends in Electronics Information & Communication Technology (RTEICT),2017
4. Kaiz Merchant, Yash Pande ,” NLP Based Latent Semantic Analysis for Legal Text Summarization”,IEEE,2018
5. Abu Kaisar Mohammad Masum, Sheikh Abujar, Md Ashraful Islam Talukder, AKM Shahariar Azad Rabby, Syed Akhter Hossain,” Abstractive method of text summarization with sequence to sequence RNNs”,ICCCNT,2019