****

**CSCE-5290**

**NATURAL LANGUAGE PROCESSING**

**Faculty:- Professor Sayed Shah**

**Project Title:- Abstractive Text Summarization using Sequence-to-sequence RNNs**

**Team Mates:**

1) Rajesh Nemani

2) Sameer Syed

3) Aishwarya Beemanapalli

4)Aravind Chennuri

**What is Text Summarization?**

The challenge of writing a succinct, precise, and fluid summary of lengthier text content is known as text summarization.

In NLP, Text summarization is the process of synthesizing large text bodies into their most prominent parts. We will use Machine Learning models for this process.

Text summarization can be used in various fields such as Law for summarizing the entire details in a better way, Interviews, Calls, YouTube videos and so on.

**Why Text summarization?**

There is a tremendous quantity of text available, and it keeps expanding daily. Consider the internet, which has a wide variety of online sites, news stories, status updates, blogs, and more. The greatest thing we can do to traverse the unstructured material is utilize search and scan the results.Much of this text material has to be condensed into concise summaries that highlight the key points so we are able to traverse it more efficiently and determine whether the longer papers actually contain the information we need.

In a book Automatic Text Summarization the author mentions few points on why do we need text summarization. They are as follows:

1. Summaries reduce reading time.
2. When researching documents, summaries make the selection process easier.
3. Automatic summarization improves the effectiveness of indexing.
4. Automatic summarization algorithms are less biased than human summarizers.
5. Personalized summaries are useful in question-answering systems as they provide personalized information.
6. Using automatic or semi-automatic summarization systems enables commercial abstract services to increase the number of texts they are able to process.

**Our Motive:**

Text summarization is very crucial now a days because of the length of data versus the time we have to process and work on it. Manually doing the text summarization is tedious, time-taking and is prone to minor errors. To overcome such tedious task, we have come up with the idea of summarizing texts using NLP techniques and models. Our motive for this project is to extract text from News papers and summarize it. In this way, we can get a simple summary of the long text and it can be helpful in future for references instead of having the burden to read the entire news which is basically lengthy.

**How do we implement and what model we use?**

In this project we are planning to do text summarization using Pytorch, with RNN model as encoder and decoder. As Recurrent neural networks (RNNs) are a form of neural network in which the results of one step are fed into the next phase's computations. RNN retains every piece of knowledge throughout time. Only the ability to remember past inputs makes it helpful for time series prediction. After loading the dataset in dataframes we will be cleaning the data so in order to remove any noises in it, then the input will be sent to encoder RNN.

**Features of Dataset:**

The dataset is of text format so the features are tokens, vectorizations, word frequency, lemmatizing

**Scope of text Segmentation**:

The majority of businesses rely on segmentation for more individualized marketing, however it is ineffective for connecting with customers on a human level.

**What is the Data?**

The data we are using is in form of .csv file which contains news tittles and news extracted from different sources.

**References**

1. <https://machinelearningmastery.com/gentle-introduction-text-summarization/>
2. <https://www.geeksforgeeks.org/introduction-to-recurrent-neural-network/>
3. <https://github.com/sunnysai12345/News_Summary>

**Github Link:**

<https://github.com/rnemani96/CSCE-5290-Natural-Language-Processing/tree/main/Project>