Problem Statement

Create a highly accurate text summary using Deep Learning algorithms

Proposed Solution

Summarization is the task of condensing a piece of text to a shorter version, reducing the size of the initial text while at the same time preserving key informational elements and the meaning of content. Since manual text summarization is a time-expensive and generally laborious task, the automatization of the task is gaining increasing popularity and therefore constitutes a strong motivation for academic research.

Automatic text summarization aims to transform lengthy documents into shortened versions, something which could be difficult and costly to undertake if done manually. Machine learning algorithms can be trained to comprehend documents and identify the sections that convey important facts and information before producing the required summarized texts.

The need for text summarization

With the present explosion of data circulating the digital space, which is mostly non-structured textual data, there is a need to develop automatic text summarization tools that allow people to get insights from them easily. Currently, we enjoy quick access to enormous amounts of information. However, most of this information is redundant, insignificant, and may not convey the intended meaning. For example, if you are looking for specific information from an online news article, you may have to dig through its content and spend a lot of time weeding out the unnecessary stuff before getting the information you want. Therefore, using automatic text summarizers capable of extracting useful information that leaves out inessential and insignificant data is becoming vital. Implementing summarization can enhance the readability of documents, reduce the time spent in researching information, and allow for more information to be fitted in a particular area.

The main types of text summarization

Broadly, there are two approaches to summarizing texts in NLP: extraction and abstraction.

Extraction-based summarization

In extraction-based summarization, a subset of words that represent the most important points is pulled from a piece of text and combined to make a summary.

Abstraction-based summarization

In abstraction-based summarization, advanced deep learning techniques are applied to paraphrase and shorten the original document, just like humans do. Therefore, abstraction performs better than extraction.

The Goal of the project

Zee Media produces hundreds of articles in multiple genres like politics, social, sports, Bollywood, health etc. The application will analyze each of these articles and produce a highly accurate text summary for these articles using Abstraction-based summarization.