

Assignment 8: Text Summarization

Instructor: Prasenjit Majumder

Learning Outcome: At the end of this assignment you will learn using extractive summarization algorithms and also evaluating summaries using ROUGE score.

1 Text Summarization

Text summarization is the technique for generating a concise and precise summary of voluminous texts while focusing on the sections that convey useful information, and without losing the overall meaning.

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.

Broadly, there are two approaches to summarizing texts in NLP

1. **Extractive Summarization:** In this approach a subset of words that represent the most important points is pulled from a piece of text and combined to make a summary.
2. **Abstractive Summarization:** In this approach new sentence are generated from original document. Sentences generated by this approach may not be present in original document.

2 Implementation

2.1 Dataset

- You are provided with the BBC business news dataset for this assignment.
- BBC business news contains news and summary in text format.

2.2 Exercise

1. Implement LexRank and TexRank summarization approaches for generating the summary for all the news articles with summary size of length 2,3 and 4 sentences.
2. For each summary generated, calculate Rouge 1 and Rouge 2 scores document wise. Finally Average out rouge scores that will be score for whole data.

3 References

1. [LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization](#)
2. <https://github.com/miso-belica/sumy>
3. <https://iq.opengenus.org/lexrank-text-summarization/>