

Text Summarization using NLP

Bhuvana Sri Likhitha Kanakam, Surya Chakra Mani Kuntha Sai Kapisetti,
Murali Venkata Ratna Sai Gunnam, Roshan Khandelwal
Under Prof. Lu Xiao

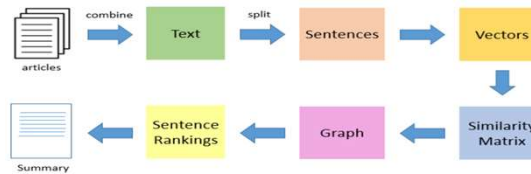
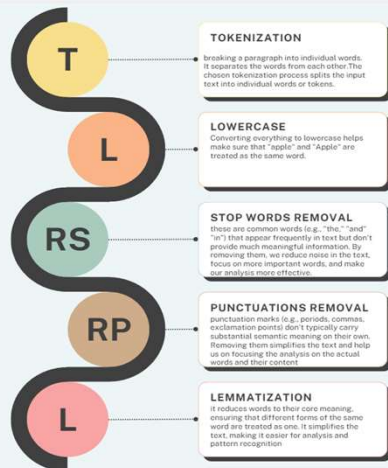
OVERVIEW

In our poster presentation, our aim is to explore Text summarization techniques which is an important aspect of Natural Language processing. With the ever-growing volume of textual information, humans feel difficult to interpret the context of a given text. To address this challenge, we explore different text summarization techniques. We are going to use two types of text summarization techniques. Extractive text summarization extracts the topmost important sentences directly from the given text as summary whereas abstractive summarization involves generating new sentences to convey the essential information.

NLP Text Summarization

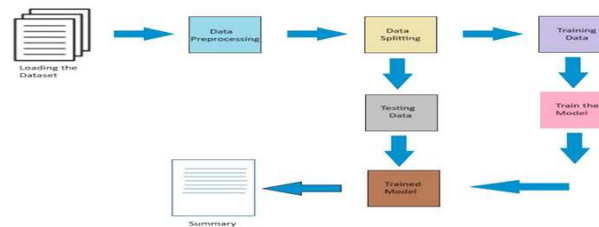


Data Pre-Processing



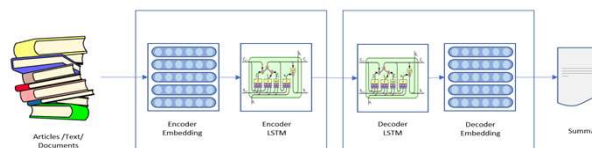
Extractive Text Summarization:

- Take some textual data with which you want to perform extractive summarization.
- Break the text into individual sentences.
- Remove unnecessary elements such as special characters, numbers, and irrelevant symbols.
- Eliminate stopwords that do not contribute much to the meaning of the text.
- Convert each sentence into a numerical vector representation.
- Compute pairwise similarity scores between sentences.
- Assign scores to sentences based on their similarity and other feature.
- Select the top-ranked sentences to construct the final summary.

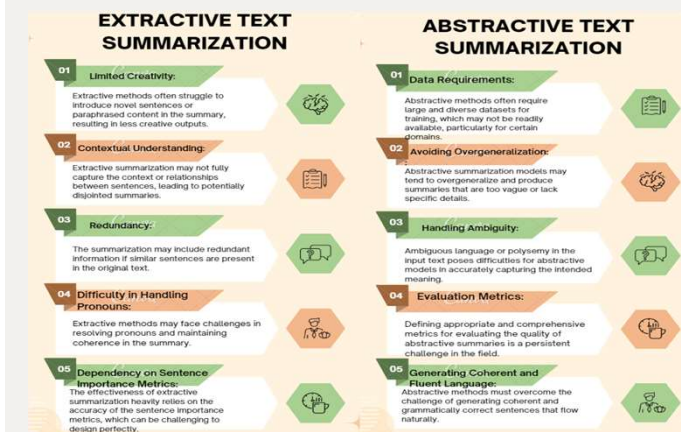


Abstractive Text Summarization:

- Take a CSV file which consists of textual data.
- Break the text into individual sentences.
- Remove unnecessary elements such as special characters, numbers, and irrelevant symbols.
- Eliminate stopwords that do not contribute much to the meaning of the text.
- Convert the entire document into a meaningful representation, often using word embeddings.
- Utilize neural network architectures, such as Long Short-Term Memory (LSTM) networks, for sequence-to-sequence mapping.
- Implement attention mechanisms to allow the model to focus on important parts of the input text while generating the summary.
- Generate a sequence of words that forms the abstractive summary.



CHALLENGES



Text summarization using Natural Language Processing (NLP) is an advanced method of extracting important information from lengthy textual data. The main goal of text summarizing is to provide brief, logical summaries that accurately convey the main ideas of the original text data. Various techniques and models are employed for text summarization, including extractive methods that select and arrange existing sentences and abstractive methods that generate novel, condensed sentences. Abstractive summarization, on the other hand, employed advanced techniques like Long Short-Term Memory (LSTM) networks to generate new, concise sentences that capture the core meaning of the content.

CONCLUSION

In summary, natural language processing (NLP) text summarization is a broad field that has the potential to greatly improve comprehension and information retrieval. Although extraction techniques are excellent at maintaining the original language and consistency, they may miss subtle meanings. Deep learning-driven abstractive techniques facilitate more inventive and adaptable summarizing; nonetheless, they present difficulties in producing precise and well contextualized content.

REFERENCES

- <https://bard.google.com/%3C0%3Ehttps://www.semanticscholar.org/paper/Ranking-Sentences-for-Extractive-Summarization-with-Narayan-Cohen/59562be2cf8e01e8b7bb7560cef56158ea171227>
- <https://link.springer.com/article/10.1007/s00521-023-08680-0>
- <https://arxiv.org/abs/1912.08777>
- <https://arxiv.org/abs/1706.06681>