

Unstructured Text Summarization – Assignment 8 (Optional)

ID: eo9232

Name: Md Reza

IE7860 – Winter 2022

Project Report

Background and Methods:

Article Summarization: This study proposes a machine learning approach to summarize unstructured text from an online article using the NLTK library. Steps include:

- Leverage **BeautifulSoup** library for web scraping.
- Tokenizing the sentences.
- Calculate the weighted frequency of each occurrence of the word.
- Calculate sentences' weighted scores.

Pre-processing:

| | |
|--|---|
| Article Transformer (Machine Learning Model) [1] | 1. Remove unwanted text 2. Remove special characters |
|--|---|

Model Evaluation:

The summarized snippet below shows that the model performance is quite promising in summarizing the article's unstructured text from the Wikipedia an online encyclopedia.

Summarized Article: Each decoder layer has an additional attention mechanism that draws information from the outputs of previous decoders, before the decoder layer draws information from the encodings. Transformers are built on these attention technologies without using an RNN structure, highlighting the fact that attention mechanisms alone can match the performance of RNNs with attention. While each attention head attends to the tokens that are relevant to each token, with multiple attention heads the model can do this for different definitions of "relevance". Transformers use an attention mechanism without an RNN, processing all tokens at the same time and calculating attention weights between them in successive layers. $q_i \cdot k_j$ (is large), this does not necessarily mean that token j will attend to token i (i.e. The decoder functions in a similar fashion to the encoder, but an additional attention mechanism is inserted which instead draws relevant information from the encodings generated by the encoders. When a sentence is passed into a transformer model, attention weights are calculated between every token simultaneously.

Summary:

Extractive Summarization and Abstractive Summarization are broadly used for text summarization. The objective of the exercise was to build an NLP model leveraging the extractive summarization approach to extract keywords and meaningful sentences from the fed article from Wikipedia and extract the meaningful summary of this article with sentences that have high weighted scores.

Reference:

[1]. Wikipedia contributors. (2022, March 22). Transformer (machine learning model). In Wikipedia, The Free Encyclopedia. Retrieved 16:41, April 21, 2022, from [https://en.wikipedia.org/w/index.php?title=Transformer_\(machine_learning_model\)&oldid=1078628228](https://en.wikipedia.org/w/index.php?title=Transformer_(machine_learning_model)&oldid=1078628228)