

***NLP Case Study*** *-*

**Medical Document Summarization with**

**pre-trained Transformers**

Name: Thomaskutty Reji

Roll No: 20122011

**Abstract**

Effective communication and understanding key points have far reaching benefits in language domains. Text summarization reduces a long-documents to main ideas and points and helps us to understand these long documents faster. Summarization of documents using modified transformers to improve long medical document summarization is performed with transfer learning with PubMed dataset. ROUGE measure will be used as the evaluation metrics. In this work pretrained transformers are used for long medical document summarization. Model metrices will be further checked for performance analysis.

**Introduction**

Text summarization helps us to easily get the relevant information from long documents. A best example of text summarization that we all are familiar with is news summarization. We have many applications which summarize the long text of news which may take much time to read.

With respect to the input type; summarization task can be classified as single document and multi-document. In terms of aim of the model the summarizations tasks are divided into Generic, where model treats all the input without bias and domain specific model add prior bias to the information. Query-based summarizations helps to answers to natural questions about the input text.[1]

Now let’s come to most discussed summarization division in NLP which is based on model output. The two variants of such summarization are, Extractive summarization and Abstractive summarization. In extractive summarization we select important words and sentences and extract them from the original text thereby create the summary. But more interestingly in abstractive summarization, we can create new summaries without using the exact sentences from the original text without losing any relevant information. Here the entire document is reproduced in a different way. It is just like we read a long article and summarizing with our own words without losing the any relevant information. Some of the key challenges with respect to text summary generation, are text identification, interpretation and text generation and analysis. The most critical phase is identifying the key phrases in the input document and using them to generate relevant information in the input document. So, it is a challenging task to generate an efficient model for document summarization.

**Objective**

The primary object of this case study is to create an efficient document summarizer which uses various transformers and other advanced NLP developments. The secondary objective includes understanding various pretrained models and their architectural designs.

**Problem statement**

Even though transformers learn much better long-term dependencies and relations, it still has limited context window problem and this affects the performance of summarization of long documents. So, I will address this problem of document summarization using various pretrained transformers with modified networks.

**Literature review**

There exist many algorithms which uses statistical techniques for making the extractive summary. Latent semantic analysis uses singular value decomposition for extractive summarization. Luhn algorithm uses tf-idf for summarization. Kl sum is another algorithm which uses word distribution similarity.

RNN based sequence to sequence modelling has attracted a lot of researchers in 2014. Recurrent Neural Networks and LSTMS were used for many NLP tasks because of their ability to model sequential data. Seq2seq models were useful in many NLP tasks including machine translation, text summarization, speech recognition and question answering system. The performance of these models was improved with attention mechanism in 2015. In 2017 Noam et al. introduced transformer, which is the sequence transduction model based on multi-headed attention and it replaced recurrent neural network for NLP tasks. [2]

The main challenges with the seq2seq models are their inability to learn long term dependency and inability to parallelize the model architecture. With transformers we are able to implement abstractive summarization and solve the problems with the seq2seq models. But transformers have its own limitations too, which involves inability to stretch beyond a certain level due to the use of fixed length context. This is also known as context fragmentation.

Language model has received many advances with the state-of-the-art pretrained models which includes GPT, BERT etc. BERT or bidirectional encoder representation from transformers combines both word and sentence representation on pretrained transformer. Now, transfer Learning is a technique which uses the pretrained model and fine tuning it for a new task. BERT framework is a new language representation model developed by google AI, uses pretraining and fine tuning to create state of art models for a wide range of tasks. These tasks include question answering system, sentiment analysis and language inferences.

In [3] Gonzalez et al. performed the summarization task in two steps namely extractive and abstractive step. Since transformers have limited size context windows, the proposed mixed extractive and abstractive approach attempts helped to solve this issue. PubMed dataset were used for the modelling and results showed that the use of pretrained transformers lead to the improvements in both extractive and abstractive steps. BERT and BART are used with first and second steps respectively. ROUGE scores were used as the evaluation metric for the model.

Liu and Lapata [4] showcased how BERT can be use with the text summarization and proposed a general framework for both extractive and abstractive summary models. They found that two staged fine-tuning approach boost the quality of the generated summaries. The proposed document level encoder tested with the three datasets and showed good performance.

Urvashi et al. [5] used pretrained decoder network where transformer Language model encodes the source and creates the summary. Experimental results on CNN/Daily Mail dataset showed that pretrained transformer language model performed better than pretrained transformer encoder-decoder network.

In [6] Torres et al. evaluated extractive text summarization with BERTSUM model using ROUGE metrices. CNN/Daily Mail dataset were used for the modelling. The proposed model scores each sentence in the document to be able to include the most relevant sentence in the summary. Abstractive summary needs to have generative capabilities and the models like BERTSUM uses fine tuning layers to add document-based context from the BERT outputs to get more efficient models like DistllBert which is having similar performance but need a lot less space and time to run.

**References**

[1] I. Awasthi, K. Gupta, P. S. Bhogal, S. S. Anand, and P. K. Soni, “Natural Language Processing (NLP) based Text Summarization-A Survey,” *Proc. 6th Int. Conf. Inven. Comput. Technol. ICICT 2021*, pp. 1310–1317, 2021, doi: 10.1109/ICICT50816.2021.9358703.

[2] A. Vaswani *et al.*, “Attention is all you need,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.

[3] L. Gonzalez, S. Lu, and W. Buchanan, “Abstractive Summarization of Long Medical Documents with Transformers.”

[4] Y. Liu and M. Lapata, “Text summarization with pretrained encoders,” *EMNLP-IJCNLP 2019 - 2019 Conf. Empir. Methods Nat. Lang. Process. 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.*, pp. 3730–3740, 2020, doi: 10.18653/v1/d19-1387.

[5] U. Khandelwal, K. Clark, D. Jurafsky, and L. Kaiser, “Sample Efficient Text Summarization Using a Single Pre-Trained Transformer,” 2019, [Online]. Available: http://arxiv.org/abs/1905.08836.

[6] S. Torres, “Evaluating Extractive Text Summarization with BERTSUM.”

[5] Web source- [https://www.machinelearningplus.com/nlp/text-summarization- approaches-nlp-example/](https://www.machinelearningplus.com/nlp/text-summarization-%20approaches-nlp-example/)