International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-12 Issue-1, October 2022

Transformer-Based Abstract Generation of Medical Case Reports

Anusha Verma Chandraju, Lydia J Gnanasigamani

Abstract*: *A medical case report gives medical researchers and healthcare providers a thorough account of the symptoms, treatment, and diagnosis of a specific patient. This clinical data is essential because they aid in diagnosing novel or uncommon illnesses, analyzing specific medical occurrences, and enhancing knowledge of current medical education. The summary of the medical case report is needed so that one can decide on further reading as going through the entire contents of a medical case report is time-consuming. In this paper, we present a deep learning methodology for the generation of the automatic summaries of the medical case reports. The final proposed fine-tuned summarizer on the test data set generated a mean precision of 0.4481 and Rouge-1 Score of 0.2803.

Keywords*: *Transformers, Healthcare, Extractive Summarization, Abstractive Summarization, Medical Research

I. INTRODUCTION

Medical case reports are extremely important in the field of medicine as they help advance our scientific knowledge of diseases and helps in better diagnosing by medical professionals. With the vast number being published every day, finding relevant medical case reports is of the utmost important. Thus, a report must have a good summary so that the reader can understand the author's intent and the key points covered so that he can make the decision for further reading. Summarization in the field of medicine has been area of little research in the past owing mainly to the lack of publicly available medical datasets due to confidentiality. The dataset [1] used for fine-tuning contains the contents of the medical case report and a section called Learning Points, which has a reworded summary of the significant parts from the text. We used this section as the corresponding summaries. Our key contribution in this paper is to generate realistic summaries of the medical case reports by employing the use of fine-tuned transformers.

In extractive summarization, the important information is captured and grouped to be presented. But, in abstractive summarization, the model rather than choosing only specific passages from the source, paraphrases the essential points of the given text using a vocabulary set that it was trained on. This ensures that when compared to extractive summarizers, the abstracts generated are less redundant and closer to what humans would write. In our paper methodology, we used both these techniques to generate the best outcome. Furthermore, we trained two state- of-the-art transformers; BART-Base and T5-Small wherein we observed that the best performance was obtained when the BART-Base model is used.

II. LITERATURE SURVEY

There generally has been lesser research into the specific field of medical summarization. Finkelstein, Joseph, et al [2] propose a novel technique to automatically identify, summarize, and score drug reactions from journal papers before presenting them to doctors in a user-friendly interface. As a result, they had produced 4.8 accurate Adverse Drug Reactions out of six different medicines. This approach was recommended to be used by doctors to quickly learn about the possible adverse drug reactions that are connected to particular medications. Gunnarsson, A [3] investigates the application of the BERT model for extractive summarization of Swedish medical data where the medical histories and summaries were generated by domain experts. This paper looked into Text Rank and two other BERT based clustering methods which are clustering-based. The implementations were assessed by employing ROUGE measures. Afantenos, Set Al [4] aim to provide a broad overview of the current state of document summary, outlining the variables on which it depends, going over assessment methods, and briefly describing the different summarization approaches. Furthermore, the characteristics of the medical industry are examined using various medical documents. Jay DeYoung et Al [5] propose MS2 (Multi-Document Summarization of Medical Studies) which is a dataset that consists of more than 470k documents and 20k summaries taken from the medical literature and test a BART-based summarization system based on this dataset and they found the preliminary findings to be encouraging. Kieuvongngam V et Al [6] address the issue of the summarization of the surge of the COVID-19 related articles due to the recent pandemic. BERT and OpenAI GPT- 2 were trained on the COVID-19 Open Research Dataset Challenge.

Manuscript received on 29 September 2022 | Revised Manuscript received on 07 October 2022 | Manuscript Accepted on 15 October 2022 | Manuscript published on 30 October 2022.
***** Correspondence Author
Anusha Verma Chandraju*, SCOPE, Vellore Institute of Technology, Vellore, India. Email: canushavema@gmail.com
Lydia J Gnanasigamani, SCOPE, Vellore Institute of Technology, Vellore, India. Email: lydiajane.g@vit.ac.in

To measure the performance, ROUGE scores and eye inspection are used to assess the outcomes. Chaves, A et Al [7] provide a comprehensive overview of the text summarizing research being done for biomedical textual data. Single-document, Generic, and Extractive summarization were found to be the most frequently used approaches, according to the data collected from the IEEE, WoS, and ACM. Machine learning techniques were used in 16 studies, and Rouge was the most used evaluation metric, being used in 26 studies.

III. PROPOSED METHODOLOGY

The content of the medical case report is first provided as input for the project's proposed deep learning methodology. The BERT extractive summarizer and abstractive summarizer are then applied to the input. Two abstractive transformer- based summarizers were specifically fine-tuned to the dataset. The medical case report's final summary is subsequently created.

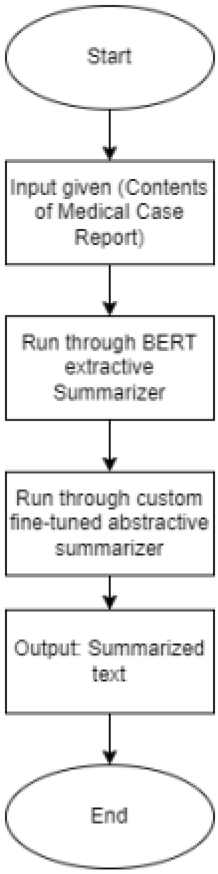

Fig. 1. Methodology

A. Dataset

The dataset consists of medical articles taken from BMJ Case Reports. It covers a wide range of subjects, including surgery, neurology, diagnostics, etc. The Training dataset for this project consists of 1000 samples with the complete contents of the medical case reports and the corresponding summaries, whereas the Test dataset, after pre-processing, consists of 497 samples. In Fig 2, The average length of the case reports in the training dataset is 8473 tokens. Fig 3 shows the visual depiction of the words in the data.

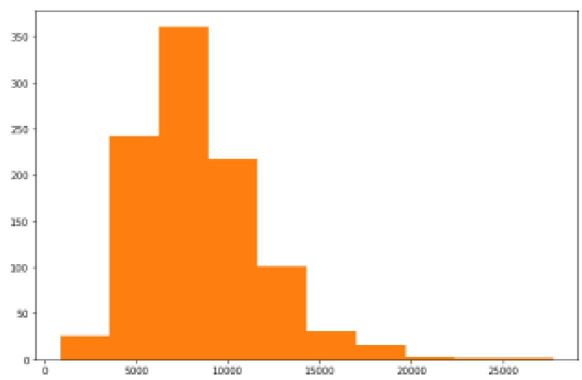

Fig. 2. Number of tokens of medical case reports

Fig. 3. Word Cloud of Training data

B. Extractive Summarizer

Here, the extractive summarizer is utilized to identify the medical case reports' most relevant sentences. This is because the input of medical case reports surpasses the maximum number of tokens that transformer-based summarizers can process. We limited the number of phrases that the summarizer can extract to six to address this problem. In this case, we employ the BERT extractive summarizer using the bert-extractive-summarizer 0.10.1 module in python, which primarily relies on the idea of embedding the sentences before using a clustering method to identify the phrases that are most closely related to the cluster's centroids. The abstractive summarizer thenreceives these processed sentences.

C. Fine-Tuned Abstractive Summarizer

The Abstractive Summarizer we have used has been fine-tuned on the 1000 data points from the train data over 3 epochs and with a total optimization steps of 750.

1) Bart-Base: A bidirectional encoder and an autoregressive decoder are combined in the transformer encoder-encoder (seq2seq) paradigm known as BART which is used as the base model for the abstractive summarizer. BART primarily works by corrupting the text input to make it accessible for pre-training (This is done using a random noise function). Then, it develops a model to recover the original text. BART particularly is well-known for excelling at comprehension tasks and in text generating scenarios such as summary and translation.

Here, we used the BART-Base model [8] which is trained on the English language.

2) *T5-Small:* Text-to-Text Transfer Transformer whichis often known as T5 is a model [9] that employs the text-to-text methodology. In contrast to BERT, every taskis transformed into a single style in this architecture, allowing it to be applied to a variety of NLP tasks. The checkpoint in our model T5-Small comprises 60 million parameters. This model has been pre-trained on the C4 dataset ("Colossal Clean Crawled Corpus).

IV.RESULTS AND DISCUSSION

For finding the best option for the abstractive summarizer, we trained two state-of-the-art transformer models on the training dataset for 3 epochs. The results can be seen in Table

1. As we can see, Bart-Base model gives us the best results when comparing the precision and F measure. Thus, we use this custom-tuned model for our abstractive summarization task.

Table- I: Comparison Between AbstractiveSummarizers

Model	Mean Rouge-1 Score	Mean F Measure
Bart-Base	0.4481	0.2803
T5-Small	0.4346	0.2465

V. CONCLUSION

With the use of this paper, we hope to give medical practitioners a tool for efficient and practical summary creation of medical case reports. At the beginning, the extractive summarizer ensures that we extract only the most crucial sentences and later, the input is then subjected to a fine- tuned abstractive summarizer for the final summary; in this example, the Bart-Base model with a mean precision of 0.4481 and Mean F Score of 0.2803 showcases the best performance. The future scope of this project includes developing an application for users to use this summarization tool. Furthermore, we are also looking at taking human feedback from domain experts on the generated summaries as a performance metric.

APPENDIX

Example 1:

Original Summary: The management of a patient with pre- existing deep venous thrombosis (DVT) who was subsequently found to have a malignancy planned for urgentlaparotomy reported.

Generated Summary: a 30 - year - old woman presented to the emergency room and was diagnosed to have a malignancy. she was treated with enoxaprin and warfarin (tablets). the patient was started on heparin (injection), intercostal (tablet) and' sachelet, but was stopped by laparotomy following a 10 – day course of intensive care forsevere pain in the right leg.

Example 2:

Original Summary: A 41 - year - old unemployed construction worker with a background of insulin - dependent diabetes mellitus, hypertension, chronic kidney disease stage 3, previous incisor tooth abscess aged 9, for which he was fitted with metal dentures, underwent a right below knee amputation. Postoperative day 8 he experienced episodes of dysphagia and vomiting postparandial. CT thorax revealed a foreign body in the midline region of the mediastinum. After two failed attempts at endoscopic extraction of the metal denture, the patient underwent surgical extraction of his metal denture.

Generated Summary: a 41 - year - old unemployed construction worker with a background of insulin - dependent diabetes mellitus, hypertension and chronic kidney disease stage 3, underwent a incisor tooth abscess. he was prone to dysphagia (haematemesis or postparandial vomiting), which was resolved by intravenous fluids with water soluble contrast study which revealed metallic dentures embedded inthe midline of the mediastinum; the patient was not aware ofhis location.

ACKNOWLEDGMENT

We would like to thank Simon Šuster and Walter Daelemans for providing us with the Clicr dataset and guiding us on the usage and acknowledge the help provided by BMJ Case reports.

REFERENCES

1. Šuster, S., & Daelemans, W. (2018). Clicr: A dataset of clinical case reports for machine reading comprehension. arXiv preprint arXiv:1803.09720. [CrossRef]
2. Finkelstein, J., Chen, Q., Adams, H., & Friedman, C. (2016). Automated summarization of publications associated with adverse drug reactions from PubMed. AMIA Summits on Translational Science Proceedings, 2016, 68.
3. Gunnarsson, A. (2022). Towards the creation of a Clinical Summarizer.
4. A fantenos, S., Karkaletsis, V., & Stamatopoulos, P. (2005). Summarization from medical documents: a survey. Artificial intelligence in medicine, 33(2), 157-177. [CrossRef]
5. DeYoung, J., Beltagy, I., van Zuylen, M., Kuehl, B., & Wang, L. L. (2021). Ms2: Multi-document summarization of medical studies. arXiv preprint arXiv:2104.06486. [CrossRef]
6. Kieuvongngam, V., Tan, B., & Niu, Y. (2020). Automatic text summarization of covid-19 medical research articles using bert and gpt-2. arXiv preprint arXiv:2006.01997.
7. Chaves, A., Kesiku, C., & Garcia-Zapirain, B. (2022). Automatic Text Summarization of Biomedical Text Data: A Systematic Review. Information, 13(8), 393. [CrossRef]
8. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... & Zettlemoyer, L. (2019). Bart Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprintarXiv:1910.13461. [CrossRef]
9. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M.,... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res., 21(140), 1-67

AUTHORS PROFILE

Anusha Verma Chandraju is a fourth- year studentat the Vellore Institute of Technology studying B.Tech in Computer Science and Engineering. Deep learning and machine learning are two of her key research areas. She has previously worked for companies offering machine learning- based solutions, including NIWE and Siemens Gamesa Renewable Energy. She is presently researching the area of natural language processing. She particularly enjoys working on initiatives that canaddress real-world problems.

Dr. Lydia J Gnanasigamani, has Completed Ph.D. from Vellore Institute of Technology. She has more than 9 years of teaching and research experience. She is a life member of CSI, and ISTE professional bodies. Currently, she is working at Vellore Institute of Technology, Vellore as an Assistant Professor in the School of Computer Science and Engineering. Her areas of interest include Data Science, Data Analytics and Deep Learning.

RetrievalNumber:100.1/ijeat.A38531012122
DOI: 10.35940/ijeat.A3853.1012122
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)