A study of examining the Contemporary Trends and Challenges in Text summarization

Kajal Agrawal

Associate Professor – MGM University Nath School of Business and Technology, Aurangabad (MS) India

Abstract:

Recent years have seen a sharp rise in the quantity of text data available from numerous sources. This extensive body of literature contains a multitude of knowledge and information that, in order to be of any use, needs to be sufficiently summarized. A lot of attention has been paid to text summarizing, however other from extractive summarizing activities like news or book summarizing, there hasn't been any real-world use for it. Text summarization produces an automatically generated summary that incorporates all relevant and important details from the source text. Textual information summarization is a difficult work for humans, and in the big data era, the rate at which information is growing makes it impractical or impossible to summarize most of it by hand. This research paper is based on secondary data sources attempt to examine the contemporary trends and challenges in text summarization. It also explores various kinds of text summarization techniques and also articulates on NLP techniques and benefits which can be leveraged for the convenient text summarization.

Keywords: Text Summarization, NLP, Multi-document, Multi-modal Summarization

I) Introduction:

Text summarization is a vital natural language processing (NLP) task that involves condensing a large body of text into a shorter, coherent version while preserving its key information and meaning. With the exponential growth of digital content across various domains such as news articles, research papers, social media posts, and legal documents, text summarization has become increasingly essential for efficiently extracting valuable insights, aiding in information retrieval, and enhancing overall comprehension. By leveraging techniques ranging from statistical methods to advanced deep learning models, text summarization algorithms strive to produce concise summaries that capture the essence of the original text, enabling users to quickly grasp its main points without having to delve into the entire

document. This introductory overview explores the significance of text summarization in the era of information overload and its diverse applications across industries and domains.

II) Objectives of the Study:

1. To present contemporary trends in Text Summarization.

2. To explore different types of Text Summarization.

3. To understand the challenges coming in the path of Text Summarization.

III) Research Methodology

This research is exclusively reliant on secondary sources of data, emphasizing an exploratory approach towards the research subject. The data is sourced from electronic journals, web portals, and textbooks, contributing to a comprehensive and in-depth exploration of the chosen topic.

IV) Literature Review:

- In a research paper based on the Recent Progress on Text Summarization, it has been understood that by shrinking the original document while preserving all pertinent information in the original. Still, computerized text summary remains a difficult task.   endeavor, and despite the several approaches that have been put out, the outcomes still differ significantly from high-quality human summaries. The majority of researchers concentrate on the method of extraction. Hence, compared to abstractive summarization, extractive summarization has a larger body of literature. This survey has examined several strategies and techniques. We get to the conclusion that, in comparison to utilising only one strategy, combining two approaches or methods is likely to yield good outcomes and improve the quality of the summary. **(Suad Alhojely & Jugal Kalita,2022)** [1]

  In An Empirical Survey on Long Document Summarization Datasets, Models and Metric, it has been deduced that comprehensive overview of long document summarization, systematically analyzed the three key components of its research settings: benchmark datasets, summarization models and evaluation metrics. It highlighted the intrinsic differences of short and long document datasets and showed that summarizing long documents requires extra compression of the source text through the identification of key narratives that are more uniformly scattered across the source documents. In this review, the researcehers a thorough analysis of long document summarising and methodically examined the benchmark datasets, summarization models, and evaluation criteria that are the three main elements of its research settings.. **(HUAN YEE KOH & et, al, 2022)** [2]

- Quantity of data at hand. Humans find it challenging to summarize lengthy texts. Because of this, automatic summarizing technologies are desperately needed in this day of excessive information. In this paper, we focused on different extraction strategies for summarizing single and multiple

documents. Some of the most popular techniques, like topic representation approaches, were explained. Machine learning, graph-based, and frequency-driven approaches. **(Mehdi Allahyari & et, al 2017)** [3]

NLP based Text Summarization Techniques for News Articles: Approaches and Challenges

- Text summary software purposefully provides a vast amount of information to aid readers in comprehending a document's or article's primary idea, regardless of language. When using the internet, people usually concentrate on the most important information, journals, movie reviews, news article highlights, and a summary of the most recent scientific discoveries. With so much information available on the internet these days, natural language processing (NLP) professionals concentrate on providing automatic summaries. The results of the recommended architecture clearly show that a serious summarization system for texts in Urdu might produce summaries that are promising. Maintaining the concept of the original text and paraphrasing it to establish connections between the many summary phrases, the evaluation findings demonstrate that summaries produced by the automatic abstractive text summarization architecture are competitive with summaries created by humans. Further research in this field might yield summaries of literature published in Urdu and other forms of information retrieval.
 **(Shafiq et al. 2023)** [4]

- With many business applications, automatic text summarization is an interesting academic topic. Summarizing can help with a range of downstream applications, such as news digests, report creation, news summarization, and headline creation, by breaking down massive amounts of information into manageable chunks. There are two sorts of summarizing algorithms that are most commonly employed. First, extractive summarizing techniques include rearranging and duplicating sections of the original text. Second, abstractive summarization techniques produce new phrases by rewording or adding terms that weren't in the original text. The majority of previous research has been extractive since abstractive summarization is challenging. The extractive approach is more convenient since it assures grammar and accuracy by copying large portions of text from the source document. Advanced abilities like paraphrase, generalization, and assimilation of real-world knowledge, on the other hand, are only possible in an abstractive framework and are required for high-quality summary. Despite the fact that abstractive summarization is a more challenging task, there has been some success thanks to recent advances in the deep learning sector. **(Sara Tarannum & et al , 2021)** [5]

- With so much information available online, there is a rising need for compressed, relevant abstracts of topics. This has led to the growth of text summarizing as a sub-branch of NLP. Accurate information facilitates more effective and efficient searching. Thus, text summary is necessary and utilized by researchers, government agencies, business analysts, marketing executives, developers, students,

and teachers. It is evident that executive summaries are necessary in order to process the necessary information in a condensed amount of time. Although abstractive summarization is slightly more complex than the extractive approach, it yields a summary that is more relevant and meaningful than the extractive method. This is because abstractive summarization involves more learning and reasoning. **(Deepali & et al 2022)** [6]

Types of Text Summarization
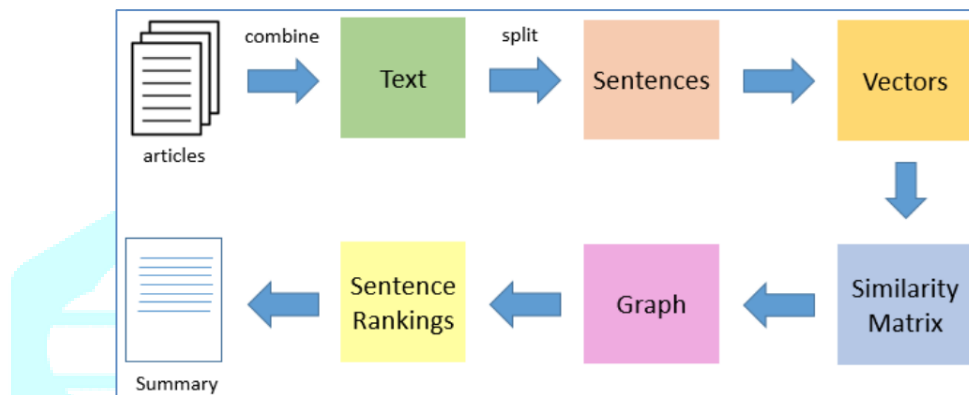
V) Text summarization uses several techniques and algorithms to give outputs. Broadly, summarization can be categorized into three types:

    a) On the basis of Input Type

    b) On the basis of Output Type

    c) On the basis of Purpose

a) On the Basis of Input type

One way of classifying Text Summarization is based upon input from the source text.

    Single Document Summarization: The input, in this case, is from a single document**.**

    Multi Document Summarization: Data is gathered from multiple documents and can be Summarising multiple documents might be difficult because it calls very extensive understanding.

b) On the Basis of Output type

the relationship between multiple sources of text and synthesize that information.

There are two types of output-based classification.

- o **Extractive:** Based on extraction In summarization, a subset of words from a document that best express its main ideas are selected and combined. We can better grasp this by visualising it as a highlighter that helps us to see the primary idea of the original text. When it comes to machine learning, Typically, extractive summarization entails weighing the crucial phrase components and using the results to create summaries.

- o **Abstractive:** Abstractive summarization does more than simply choose words and phrases from the original text, in contrast to extractive summarization. Instead, it creates a summary that is a distinct representation of the original content by utilising methods like Natural Language Generation (NLG). After comprehending the text's content and context, it creates the summary. It calls for a deeper comprehension of the subject matter and the capacity to produce fresh writing without distorting the meaning of the original data.

c) On the Basis of Purpose

- ✓ **Generic**: Generic summarizers make no assumptions about the content of the text. They simply summarize the original source content

- ✓ **Domain-specific:** In this kind, domain expertise is used to produce a summary that is more accurate. Research papers from a particular field, for instance, can be used as source text. They can use the original materials to construct the summary by combining particular language, context, and knowledge.

- ✓ **Query-based**: This kind of summarization mostly focuses on natural language queries. It summaries publications that address complicated issues related to the provided data, just like search engine results do. For instance, when we use Google's search engine to ask a question, it displays articles or web pages that provide the answers.

VI) Benefits of NLP Text Summarization

NLP text summarization offers several benefits in various applications, including:

Saving Time and Effort

Time and effort can be saved by having NLP Text Summarization automatically produce succinct summaries of large documents. This is particularly helpful in scenarios where processing a lot of text is required.

Improved Information Retrieval: Users can more easily browse and extract pertinent information from a huge body of papers by using summarized texts as abstracts or snippets. By giving consumers succinct and insightful content summaries, this can improve information retrieval activities like document search, recommendation systems, or information extraction.

Scalability and Automation: NLP Text Summarization is scalable and automatically applicable to massive amounts of text. This makes it appropriate for real-time processing of large amounts of data, as in online forums, social media feeds, and news websites, where human summarization would be unfeasible and time-consuming.

Extraction of Key Information: Information extraction jobs can benefit from NLP Text Summarization's ability to extract and highlight relevant concepts, keywords, or entities from the original text.

Customization and Personalization: Additionally, it can be tailored and adapted to particular fields, business sectors, or user preferences. This enables customized summarization outcomes that meet particular needs and can be applied to news aggregation, content curation, or customized information retrieval, among other uses.

Preserving Semantic Meaning: One of the primary challenges in text summarization is ensuring that the summary accurately captures the semantic meaning and nuances of the original text. Maintaining coherence and coherence while condensing the content can be difficult, especially when dealing with complex sentences or ambiguous language.

VII) Challenges in Text Summarization

1) Handling Abstractive Summarization

Abstractive summarization involves generating summaries in natural language, which requires the model to understand the text deeply and generate novel phrases. This is challenging as it involves generating language beyond what is explicitly stated in the source text, often requiring a deep understanding of context and domain-specific knowledge.

2) Dealing with Domain Specificity

Text summarization models may struggle when faced with texts from highly specialized domains such as legal documents, medical literature, or scientific research papers. Understanding domain-specific terminology and context is crucial for producing accurate summaries in such cases.

3) Addressing Information Redundancy

Many texts contain redundant information or repetitions, which can lead to summaries that are longer than necessary or fail to capture the key points effectively. Identifying and eliminating redundant information while retaining important details is a significant challenge in text summarization.

4) Addressing Bias and Fairness

Text summarization models are susceptible to biases present in the training data, which can lead to biased or unfair summaries, especially in sensitive topics such as politics, gender, or race. Ensuring fairness and mitigating biases in the summarization process is crucial for producing balanced and objective summaries.

VIII) Conclusion

Through a comprehensive analysis of model architectures, mechanisms, and experimental procedures, the paper attempted to provide insights into the diverse approaches utilized in summarization models to generate concise summaries of lengthy documents. Moreover, it also shed light on the scarcity of easily applicable evaluation metrics tailored for long document summarization, emphasizing the ongoing need for advancements in this area. Drawing from diverse perspectives presented in recent research studies, it explored the practical applications of long document summarization models across various domains, underscoring their significance in facilitating efficient information processing for tasks such as news digests, report creation, and information retrieval.

References:

1. Suad Alhojely Jugal Kalita   ACM Comput. Surv., Vol. 1, No. 1, Article 1. Publication date: January 2022. International Conference on Computational Science and Computational Intelligence (CSCI)

2. HUAN YEE KOH, JIAXIN JU, MING LIU∗, SHIRUI PAN∗ ACM Comput. Surv., Vol. 1, No. 1, Article 1. Publication date: January 2022.

3. Mehdi Allahyari_, Seyedamin Pouriyehy, Mehdi Assefiy, Saeid Safaeiy, Elizabeth D. Trippez, Juan B. Gutierrezz, Krys Kochuty (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10, 2017

4. Shafiq et al. (2023), PeerJ Comput. Sci., DOI 10.7717/peerj-cs.1176

5. Sara Tarannum, Piyush Sonar, Aashi Agrawal Krishnai Khairna International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Vol 8 Issue 12

6. Deepali K. Gaikwad1 and C. Namrata Mahender2 International Journal of Advanced Research in Computer and Communication Engineering   A Review Paper on Text Summarization Vol. 5, Issue 3, March 2016

7. Suleiman D, Awajan A. 2020. Deep learning based abstractive text summarization: approaches, datasets, evaluation measures, and challenges. Mathematical Problems
   in Engineering 2020:9365340 DOI 10.1155/2020/9365340.

8. Y. Sankarasubramaniam, K. Ramanathan, and S. Ghosh, ″Text summarization using wikipedia,″
   Information Processing & Management,

9. www.topcoder.com