(REVIEW ARTICLE)

Automatic text summarization–A systematic literature review

Veena R [1, *], D. Ramesh [2] and Hanumantappa M [3]

[1] Department of Computer Science, Seshadripuram College, Bengaluru, Karnataka, India.
[2] Department of Computer Application, SSIT, Tumkur, Karnataka, India.
[3] Department of Computer Application, Bangalore University, Bengaluru, Karnataka, India.

Abstract

Automatic summarization is the act of computationally condensing a set of data to produce a subset (a summary) that captures the key ideas or information within the original text. To do this, artificial intelligence algorithms that are tailored for diverse sorts of data are frequently created and used. Ten research articles considering databases like IEEE, Scopus, and Springer Nature have been considered. The paradigm shift that AI has created in the field of Automatic Text Summarization is discussed in detail.

Keywords: Automatic summarization; Keyword Extraction; Abstraction; Natural Language Processing; Dataset

1. Introduction

Text summarization automatically creates a summary that contains key phrases and all pertinent data from the original document. When seen from the summary results, extractive and abstractive methods are among the main ones. A real-time summary and abstractive summation are now the focus of research as an extractive summary nears maturity. There are not many papers that can give a comprehensive picture of the present state of research on this topic, despite the fact that there have been numerous successes in the acquisition of datasets, approaches, and techniques published. The present research work gives an extensive review of 10 articles pertaining to Automatic text summarization.

2. Literature review

Avinash Payak et al. have researched Automatic Text Summarization and Keyword Extraction using Natural Language Processing. People spend a lot of time searching the internet for pertinent information as part of the time-consuming process of gaining and assimilation knowledge from many sources. The authors have stated that the major goal of the suggested system is to scrape data from websites and provide users with the option to choose the website of their choice while also offering a summary and keywords from the information taken from multiple websites [1].

K Usha Manjari et al. have researched Extractive Text Summarization from Web pages using Selenium and TF-IDF algorithm. The authors have stated that it takes a lot of time to get an overview of the information included in various documents. Similarly, to this, finding specific information online while scouring numerous websites and web pages is a tedious effort. One of the methods that are currently used the most frequently to obtain a succinct and brief summary of the material is automatic text summarising. In this research, a novel method for gathering data from various websites on the internet is proposed to produce an extracted summary of the information based on the user's query. Selenium-based web scraping is also covered. For text summarization, the Term Frequency-Inverse Document Frequency (TF-

* Corresponding author: Veena R

IDF) method is used. The suggested method is distinctive and effective for producing summaries in response to user requests [2].

J.N. Madhuri et al. have researched extensively on extractive text summarization using sentence ranking. The method of extracting the most important and relevant information from a text is known as text summarising. It employs two strategies. There are two types of text summarization: abstractive and extractive. An extractive text summarization refers to the process of removing significant text from a text file or source manuscript. The research presents a novel statistical technique for extracting text summarization on a single document. Sentence extraction is a method that presents the idea of the supplied text in a condensed form. The weights that are assigned to sentences determine how they are ranked. Important sentences are retrieved from the input material using highly scored sentences, leading to an excellent summary of the input content [3].

Alexios Gidiotis et al. have researched A Divide-and-Conquer Approach to the Summarization of Long Documents. The authors have suggested a new divide-and-conquer strategy for summarising lengthy documents. This strategy decomposes the issue into a collection of more manageable summarising issues by taking advantage of the discourse structure of the document and sentence similarity. The authors have specifically suggested dividing a lengthy document and its summary into numerous source-target pairs so that a model may be trained to sum up each section of the document separately. The final full summary is created by combining these partial summaries. By using this method, we can break down the difficult task of summarising lengthy documents into smaller, easier tasks, which reduces computing complexity and increases the number of training examples while also reducing noise in the final summaries compared to the conventional method [4].

Sakdipat Ontoum et al. have researched on Automatic Text Summarization of COVID-19 Scientific Research Topics. Through the identification and extraction of pertinent information from articles, automated text summarization benefits the scientific and medical fields. Automatic text summarising is a technique for condensing text files so that readers may quickly access pertinent information in the original text. Several recent summarising research articles make use of deep learning techniques, are analysed after which it is explained how they relate to the COVID-19 research publications. The readability test measures how easily a reader can comprehend written content. In natural language processing, a text's readability is determined by its content [5].

Pratik K. Biswas et al. have researched Extractive summarization. One of the most difficult and intriguing issues in natural language processing is automatic text summarization (NLP). The practice of taking the most crucial information from a text and succinctly presenting it in fewer sentences is known as text summarising. A call transcript is a written account of a phone call between a customer (the caller) and an agent(s) (customer representative). The majority of open-source automatic text summarizers were designed to summarise continuous materials like articles and stories, not called transcripts, which present specific difficulties that are not well addressed by these tools [6]

Asmaa Elsaid et al. have researched A Comprehensive Review of Arabic Text Summarization. How we gather, assess, and comprehend data has evolved as a result of the expansion of internet and offline data. Large text documents are sometimes challenging to understand and time-consuming to extract important information from. The aforementioned issues are addressed by text summarising approaches, which condense lengthy texts while preserving their key information. These strategies depend on providing their users with filtered, high-quality content quickly. Automated text summary of vast amounts of data is difficult due to the enormous amounts of data generated by technology and numerous sources. Automatic text summarising methods can be divided into three categories: extractive, abstractive, and hybrid. Despite these earlier methods, the generated summaries still fall well short of the summarization created by human specialists [7].

Aakanksha Sharaff et al. have researched a Feature-based cluster ranking approach for single document summarization. Text summarization is a method for distilling a big collection of documents. It produces a summary that succinctly and accurately conveys the overarching information found in lengthy text sources. This study presents a model for producing a single document text summary. Based on extractive summarization, this model. By applying the similarity measure technique, the suggested work extracts the informative aspects and creates the scoring of sentences. Sentence clusters are created after the sentence scores have been generated. Sentences from each cluster that received high rankings in terms of their relative value are included in the final summary. Clusters and the sentences within each cluster are ranked. The most important sentences from a text document are used to construct a summary of the text[8]

Batuhan Baykara et al. have researched Abstractive text summarization and new large-scale datasets for agglutinative languages Turkish and Hungarian. Accessing the key information that is pertinent to a user's needs is becoming more important as a result of the exponential growth in the amount of documents on the Web, which raises the profile of text

summaries. Recent advancements in deep learning have redirected text summarization research away from extractive techniques and towards more abstract techniques. Other languages cannot advance because the majority of research and resources are still only available in English. Low-resource languages require the collection of extensive resources suited for such activities. In this study, we make available two sizable datasets (TR-News and HU-News) that can be used as benchmarks for Turkish and Hungarian abstractive summarization tasks [9].

Arun Kumar Yadav et al. have researched Extractive text summarization using a deep learning approach. These days, social media and the Web are awash in vast amounts of unstructured data. People struggle to rapidly locate pertinent information in such a huge unstructured corpus. Text summarization is the process of gathering pertinent information and condensing it without distorting its original meaning. Manual text summarising is time-consuming, expensive, and labour-intensive to complete. In the past, researchers tried a variety of machine-learning techniques to summarise the text, but the results were still lacking [10].

3. Research Methodology

The research methodology consisted of two major steps namely

- Collection of research articles
- Analysis of research articles

3.1. Collection of research articles

The research articles on applications of artificial intelligence were collected from databases of IEEE, Scopus, and Springer Nature. Totally ten research articles were collected. The research articles on implementing Automatic Text Summarization using various deep learning, and machine learning techniques are considered.

3.2. Analysis of research articles

From the analysis of 10 research articles the following points are observed: Extractive and abstractive text summarising techniques are the two main categories. Extractive summarization is the process of choosing specific sentences from a text to serve as its summary. Due to its 1950s beginnings, extractive summarization techniques have been widely used for quite some time. Instead of focusing on trying to understand the text's content, it is more important to study the significance of each sentence and how they relate to one another. On the other hand, abstractive summarization focuses on attempting to comprehend the text's meaning and then delivering a summary based on that, which may or may not include the same lines as those in the original text. In a move towards more human-like summaries, abstractive summarization strives to construct its own sentences.

4. Conclusion

Automatic Text Summarization holds the key to a wonderful future where we will all be able to make better judgments from data.

Compliance with ethical standards

Disclosure of conflict of interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this article.

References

[1]   Payak, A., Rai, S., Shrivastava, K., & Gulwani, R. (2020, July). Automatic text summarization and keyword extraction using natural language processing. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 98-103). IEEE.

[2]   Manjari, K. U., Rousha, S., Sumanth, D., & Devi, J. S. (2020, June). Extractive Text Summarization from Web pages using Selenium and TF-IDF algorithm. In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184) (pp. 648- 652). IEEE.

[3]   Madhuri, J. N., & Kumar, R. G. (2019, March). Extractive text summarization using sentence ranking. In 2019 international conference on data science and communication (IconDSC) (pp. 1-3). IEEE.

[4]   Gidiotis, A., & Tsoumakas, G. (2020). A divide-and-conquer approach to the summarization of long documents. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 3029-3040.

[5]   Ontoum, S., & Chan, J. H. (2022, August). Automatic Text Summarization of COVID-19 Scientific Research Topics Using Pre-trained Models from Hugging Face. In 2022 Research, Invention, and Innovation Congress: Innovative Electricals and Electronics (RI2C) (pp. 1-8). IEEE.

[6]   Biswas, P. K., & Iakubovich, A. (2022). Extractive summarization of call transcripts. IEEE Access, 10, 119826-119840.

[7]   Elsaid, A., Mohammed, A., Fattouh, L., & Sakre, M. (2022). A Comprehensive Review of Arabic Text summarization. IEEE Access.

[8]   Sharaff, A., Jain, M., & Modugula, G. (2022). Feature-based cluster ranking approach for single document summarization. International Journal of Information Technology, 14(4), 2057-2065.

[9]   Baykara, B., & Güngör, T. (2022). Abstractive text summarization and new large-scale datasets for agglutinative languages Turkish and Hungarian. Language Resources and Evaluation, 56(3), 973-1007.

[10]   Yadav, A. K., Singh, A., Dhiman, M., Kaundal, R., Verma, A., & Yadav, D. (2022). Extractive text summarization using deep learning approach. International Journal ofInformation Technology, 14(5), 2407-2415