An Analytical Study of Text Summarization Techniques

Bavrabi Ghosh[1], Aritra Ghosh[2], Subhojit Ghosh[3], and Anupam Mondal[4]

Institute of Engineering & Management, Kolkata, India,
University of Engineering & Management, Kolkata, India
bavrabi.ghosh@iem.edu.in, aritra.ghosh2021@iem.edu.in,
subhojit.ghosh2021@iem.edu.in, anupam.mondal@iem.edu.in

Abstract. Text summarization is the process of condensing lengthy texts into concise and coherent summaries, capturing the main points of the document. It presents a significant challenge in machine learning and natural language processing (NLP) due to the vast volume of digital data available. There is a growing demand for algorithms capable of automatically condensing extensive texts into accurate and understandable summaries to effectively convey the intended message. Machine learning models are typically trained to comprehend documents, extracting essential information to produce the desired summarized output. The application of text summarization offers several advantages, including reduced reading time, accelerated information retrieval, and efficient storage of more information. In NLP, two primary methods for text summarization exist: extractive and abstractive. The extractive approach identifies key phrases within the source document and assembles them to form a summary without altering the text's content. The abstractive technique paraphrases and condenses sections of the source document. In deep learning applications, abstractive summarization can overcome grammar inconsistency issues often encountered in the extractive method. Our analysis has examined various existing methods for text summarization, including unsupervised, supervised, semantic, and structure-based approaches, and has critically assessed their potential and limitations. Specific challenges highlighted include anaphora and cataphora problems, interpretability issues, and readability concerns for long texts.To address these challenges, we propose solutions aimed at improving the quality of the dataset by addressing outliers through the integration of corrected values obtained from human-generated inputs. As research in this domain progresses, we anticipate the emergence of innovative breakthroughs that will contribute to the seamless and accurate summarization of lengthy textual documents.

Keywords: Summarization, Text Summary, Automatic Text Summarization, Natural Language Processing, Abstractive Method, Extractive Method, Deep learning Approach, Unsupervised Method, Supervised Method, Anaphora Problem, Cataphora Problem, Semantic Approach, Structure Based Approach, Machine Learning, Readability Concern

1 Introduction

Summarization is the process of separating the key bits from a larger piece of material while retaining the core ideas and concepts. It is required in a variety of settings. In the world of journalism, news pieces are frequently abbreviated to highlight the most important aspects of an event, allowing readers to stay informed despite their hectic schedules. It allows researchers to quickly comprehend the important findings and techniques of relevant studies, allowing for a more efficient examination of the current literature. Long business reports are synthesized into simple summaries in the corporate environment for executives who need a quick overview before making critical choices. Summarization algorithms are used by social media platforms to give users condensed updates that capture the substance of discussions.

Text condensation techniques rely on diverse methodologies to distill key ideas and concepts from larger textual content. The fuzzy logic approach leverages fuzzy set theory and fuzzy logic principles to handle imprecise or ambiguous information present in texts. Concept-driven methods aim to pinpoint and extract pivotal notions from the text, generating summaries based on the identified concepts and their interrelationships. Latent-semantic techniques, such as latent semantic analysis (LSA) or latent Dirichlet allocation (LDA), seek to unveil the underlying semantic patterns inherent in the text. Machine learning algorithms, including Naive Bayes, Decision Trees, or Support Vector Machines, are trained on labeled datasets to discern patterns conducive to summarization. Neural network architectures, encompassing deep learning models like Recurrent Neural Networks (RNNs) or Transformers, are harnessed to learn representations and generate summaries. Conditional Random Fields treat summarization as a sequence labeling task, employing CRFs to identify and extract salient sentences or phrases. Tree-based approaches, such as Tree Summarization or Rhetorical Structure Theory, scrutinize the discourse structure of the text to pinpoint pertinent information. Template-driven methods rely on predefined templates or schemas to extract relevant details from the text and generate summaries based on the populated templates. Rule-based techniques apply manually crafted rules or heuristics to the text to identify and extract crucial information for summarization. The ontology method capitalizes on domain-specific ontologies or knowledge bases to prioritize essential concepts and information. Multimodel semantic approaches synthesize multiple semantic representations, such as concepts, entities, and relations, to generate summaries. Information item methods identify and rank key information units or items based on their significance, utilizing the top-ranked items to construct the summary. Semantic graph-based techniques construct a semantic graph depicting the relationships between entities, concepts, and events in the text, which is then utilized for summarization.

While the aforementioned text summarization methods have their merits, they possess limitations such as over-simplification, dataset biases, reliance on static knowledge bases, and a focus on extractive rather than abstractive summaries. To overcome these limitations, we have curated a custom dataset capturing complexities across diverse text sources. Leveraging this dataset, we are tainning a novel neural network model combining attention mechanisms, graph neural networks, and transfer learning. Our model aims to generate high-quality abstractive summaries by accounting for context, semantic relationships, and domain knowledge, thereby advancing the state-of-the-art in robust and generalizable text summarization.

Oftentimes, the algorithms used by summarization models are not sufficient to find the optimal result the user is looking for. The main reason might be the use of a single template for summarization that often fails to capture the detail and distinctness of the document. The user may have to input particular details or summarize the paper in segments to get the relevant, useful information he is looking for. Other problems include a lack of model accuracy. It may be necessary to spend a lot of time comprehending complex topics before even attempting to condense them because creating a summary demands a comprehensive mastery of the original material.

All these factors combined motivated us to ease the research work of our fellow researchers and make an all-inclusive comprehensive document on summarization.

The overall structure of the paper is mentioned below. The history of tachygraphy is written in Section 2. Thereafter, we have shown different types of summarization methods in Section 3. Challenges, important features, and a brief discussion have been described in Sections 4, 5, and 6. Finally, Section 7 contains the future scopes and closing remarks.

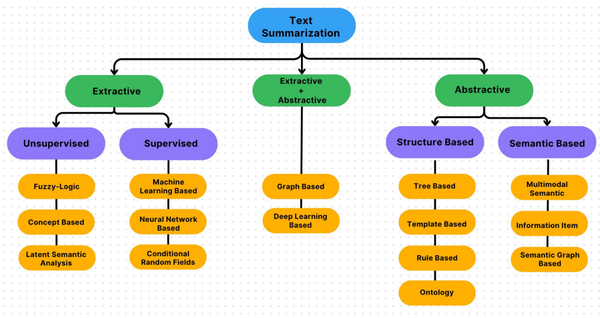

Fig. 1: Conventional methods

2 Background Study

The subsequent section delves into various approaches and techniques employed in text summarization, culminating in a comprehensive overview of the document. We conducted an exhaustive examination of the current landscape of summarization scopes and methodologies. This was undertaken with the aim of gaining insight into the complexities associated with existing methods and identifying the factors contributing to the suboptimal performance of certain approaches.

Upon conducting a survey analysis of the existing methods of summarization across various research papers it was found that a lot of summarization models are still in use. Historically, abstract summarization and extraction summarization are the two most common methods. Extractive summarization in its essence uses a ranking algorithm that produces the most occurring words from the content exactly as they appear [6]. How they work is that they calculate the frequency of each word in the document and rank them in descending order, taking only the first 'n' into consideration, for the sake of simplicity let us assume 'n' is ten here. Then they rank the sentences based on how many such important words the sentence contains. If a sentence has all 10 of these words, which generally is not the case, but let us assume an ideal situation, then that sentence will be of the highest priority. Let us assume the summarized document will have 'm' such sentences, here 'm' is selected based on how much we want to compress the document. Then the sentences are shown as they appear. Abstractive summarization initially follows the same principle but the place where it differs is that instead of exactly displaying the sentences it tries to formulate new sentences by incorporating the meaning it understood [5].

Now let us dive into some of the technicalities of each of these methods. In 1969, an automatic text summarization system was developed that went beyond the conventional keyword-based approach, which relied on frequency-dependent weights [8]. This pioneering system incorporated three additional methods to calculate the weights of sentences:

1. Cueing Method - This is predicated on the supposition that the existence of when specific cue words are absent in the dictionary determines the importance of a phrase.
2. Tile Method - This method calculates sentence weight based on all content words found in the text's title and subheadings.
3. The Location Method is founded on the premise that sentences appearing at the beginning of both the text and a paragraph are more likely to carry significance.

The Trainable Document Summarizer of 1995 was designed to perform sentence extraction tasks using a variety of weighting heuristics [8]. There are several features available in this feature set, including Fixed Phrases, Sentence Length Cut-Offs, Paragraphs, Uppercase Words, and Thematic Words.

3 Summarization Methods

3.1 Extractive Method:

1. **Term Frequency-Inverse Document Frequency method:** This is also known as TF-IDF for short. On using inverse sentence-frequency and conventional weighted termfrequency approach, where it counts the sentence numbers containing a specific term, a sentence-level bag-of-words model is constructed. After that, phrases with only the highest ratings are included in the summary. These sentence vectors are assessed based on their similarity to the query, effectively applying the Information Retrieval paradigm directly to the summarization process [7].
   Summarization is typically tailored to specific queries, but it can be adapted to a more generic form. Frequently occurring non-stop words within the document(s) can serve as query words to generate a generic summary. These phrases yield general summaries as they encapsulate the essence of the text. In the context of sentence term frequency, it often remains at 0 or 1, as a single content word doesn't usually appear frequently within one sentence. The transition from query-based summary generation to generic summarization occurs when users formulate query words similarly to how they do in information retrieval.
2. **Cluster Method:** Documents are typically structured to cover various themes in an organized sequence. Consequently, it's logical to expect that summaries would encompass these diverse "topics" present in the papers. Some summarization methods take into account this factor by utilizing clustering techniques. Term frequency inverse document frequency scores are employed representing words within the documents. The term frequency within the context of these clusters represent the mean number of occurrences per document. IDF values are calculated using the corpus. The summarization process inputs clustered documents, where one

cluster represents one distinct topic. Each topic can be illustrated as the words highest in term frequency-inverse document frequency (TF-IDF) grades within the cluster are selected.
For a sentence to be selected it is based on several criteria. Firstly the sentences are chosen based on their similarity to the central idea of cluster Ca. Secondly, the position of the sentence within the document denoted as La, is considered. Lastly, the similarity of a sentence to the sentence at first in the document referred to as Fa, is taken into account. A sentence's overall score, denoted as Si, is calculated as the weighted average of these three variables:

$$Si = (W1*Ci) + (W2*Fi) + (W3*LiSi) = (W1*Ci) + (W2*Fi) + (W3*Li)$$

3. **Graph Based Approach:** In the context of the graph-theoretic approach, sentences within documents are transformed into nodes within an undirected network, following standard preprocessing techniques. Each sentence is represented as a unique node. When two sentences share specific common terms or their similarity (e.g., cosine similarity) surpasses a predefined threshold, they are connected by an edge in the network. Selected sentences from the relevant sub-graph are used exclusively for query-specific summaries. The sub-graphs can be selected for general summaries, in contrast.
Another important outcome of graph-theoretic analysis is the identification of key sentences within a document. In this context, significant sentences within the partition are nodes that are highly connective (i.e., a great number of edges connecting to that node). These highly connected nodes are accorded greater priority for inclusion in the summary [1].
4. **Latent Semantic Analysis Method:** In the LSA Method the singular value decomposition is an exceptionally efficient mathematical technique for identifying the principal orthogonal dimensions within multidimensional data. Even if the documents do not include the same exact phrases, SVD is particularly good at grouping them together when they are semantically similar. When words appear in the same singular vectors, they are connected together because they frequently occur in comparable situations. As a result, topic words and content sentences can be extracted from documents using SVD, contributing to a deeper understanding of the data's structure and semantics.

Coming at our second approach we have a slightly advanced approach of Abstractive Summarization. Abstractive summarization is a text summarization technique where a machine generates a concise summary of a longer document by interpreting and rephrasing the content in a human-like manner. Abstractive summarising uses natural language creation to provide summaries that may include innovative words and phrases, as opposed to the extractive summary, which chooses sentences verbatim from the original text. This method is more adaptable and 4 creative but also more difficult to execute effectively because it seeks to capture the core meaning and context of the source text. Let us take a look at the most popular methods of abstractive summarization.

3.2 Abstractive Summarization:

1. **Seq2Seq (Sequence-to-Sequence) Models:** Seq2Seq Models take a variable-length input sequence, such as a sentence, and use an encoder to convert it into a fixed-length context vector or hidden representation. This context vector encapsulates the pertinent details from the input sequence. The decoder then utilizes this context vector along with a "start token" to initiate [4] the process of generating the output sequence step-by-step. Attention mechanisms can be employed to focus on specific segments of the input sequence. The decoder continues generating elements until an "end token" is produced or a predefined maximum length is reached. [17].
2. **Pointer-Generator Networks:** Pointer-Generator Networks also use an encoder-decoder setup, with the encoder processing the input document and the decoder generating the summary. A key feature is their ability to selectively copy words from the source text via a copy mechanism. The model calculates probabilities for generating each word from its vocabulary versus copying it from the input. It leverages an attention component to identify which input words should potentially be copied. By combining the generation and copy probabilities, the model can flexibly pick vocabulary words or copied words for the summary output. Coverage vectors track which input words have already been attended to, preventing excessive repetition and promoting diversity in the final summary. [15].

3. **BERTSUM:** BERTSUM first tokenizes the input document into word pieces and encodes them into contextualized embeddings using a pre-trained BERT model. It computes the salience scores for each sentence in the document. This is done by applying a feedforward neural network to the BERT embeddings of each sentence, capturing their importance. The model employs an intra-sentence and inter-sentence scoring mechanism to refine the sentence importance scores. Intra-sentence scoring assesses the importance of each word within a sentence, while inter-sentence scoring measures the importance of sentences relative to each other. BERTSUM selects sentences that have greatest important scores, effectively identifying the absolute prominent sentences in the document. Selected sentences are compressed by removing less relevant words and retaining the most informative ones. The compressed sentences are concatenated to form the summary. BERTSUM does not generate abstract summaries; it extracts and combines important sentences.

3.3 Deep Learning Summarization:

1. Bidirectional and Auto-Regressive Transformers (Abbreviation - BART): BART is a model which is transformer based and has been purposefully designed for sequence-to-sequence tasks, with a particular focus on summarization. BART has demonstrated its capability to achieve state-of-the-art performance in abstractive summarization tasks.
2. Generative Pre-trained Transformer (Abbreviation - GPT): GPT versions, including GPT-2 and GPT-3, can be fine-tuned for abstractive summarization. They generate summaries by predicting words or phrases that capture the essential information from the source text.
3. Text-To-Text Transfer Transformer (Abbreviation - T5): T5, another transformer-based model, adopts a unique approach by treating summarization as a text-to-text problem. In this framework, the input text is transformed into a summary text. T5 offers the flexibility of being fine-tuned for a wide range of summarization tasks, making it a versatile and powerful tool for various summarization applications.
4. Pegasus: An abstractive model based on transformers. It combines pre-training with fine-tuning and has shown strong performance in various summarization benchmarks.

Finally, we summarize the results of the papers for the reader's convenience in Table 1.

4 Challenges

The primary objective of any ATS (Automatic Text Summarization) system should be to generate summaries that closely resemble human-generated summaries. However, achieving this goal poses significant challenges for existing ATS systems. These challenges include:

1. **Evaluation:** Assessing the quality of automatic text summaries is a complex task. Different datasets and metrics can yield varying results and might favor specific summarization techniques. While common datasets and metrics can produce satisfactory results, they come with their own issues. Metrics like precision and recall can be misleading and might not effectively evaluate sentences with semantic or syntactic errors. This can lead to high scores for unimportant sentences while overlooking grammatically incorrect yet meaningful ones.
2. **Important Sentence Selection:** Identifying the most crucial sentences in a text is subjective. Standardizing the selection process according to benchmarks can affect the resulting summary. Incorporating user-specific data can help address this challenge in professional summarization.
3. **Anaphora Problem:** Replacement of subjects with synonyms and pronouns is a common challenge in text summarization. Addressing this problem entails the identification of which pronoun corresponds to a specific word, a task that can pose considerable complexity for machine-based systems [13].
4. **Predefined Template:** While natural language processing has made remarkable progress in ATS, these methods often rely on predefined templates for summarization tasks. They cannot generate entirely new sentences independently, necessitating the use of specific templates.
5. **Long Sentences and Jargon:** Current text summarization models excel at summarizing shorter sentences but may struggle with longer sentences and specialized jargon. Addressing this limitation requires the development of architectures capable of effectively summarizing longer sentences and handling domain-specific terminology.
6. **Interpretability:** Abstractive models can have trouble expressing emotions in written text and preserving the subtleties of human language, even though their main purpose is to produce succinct representations of the original information. The intricacy of human language and its emotional components make it difficult to achieve interpretability with abstractive models.

Table-1: An overview of different summarization methods

Algorithms			**Limitations**
Extractive	**Unsupervised**	Fuzzy logic	Post-processing should remove redundancies to improve the quality of summarization.
		Concept-based	The summary should use similarity measures to reduce redundancy, which can affect quality.
		Latent-Semantic	LSA-generated summaries take a long time.
	Supervised	Machine Learning	To make good summaries, it has to be trained and improved on a large set of data.
		Neural Network	Both the training phase and the application phase are quite slow with neural networks. Training data also requires human interruption.
		Conditional Random Fields	Linguistic features are not taken into account in the use of CRF. It also needs an external domain specific corpus.
Abstractive	**Structural**	Trees	The text ignores context and important phrases in the text, resulting in a failure to recognize the relationships between sentences. Another issue is that it consistently emphasises syntax rather than meaning.
		Template based	As the templates are pre-defined using this technique, the summaries lack variation.
		Based on Rules	It takes a long time to create regulations. It is also difficult to manually write the rules.
		Ontology method	The process of creating a suitable ontology is time-consuming and limited to a single domain.
	Semantic	Multi-model semantic	The framework must be automatically analysed because humans now manually evaluate it.
		Information item	Generating grammatical and meaningful sentences from the material is difficult. The linguistic quality of summaries is low due to incorrect parses.
		Based on Semantic Graph	Limited to single document abstractive summarization.

7. **Cataphora Problem:** Word ambiguity from various contexts or meanings can affect how well sentences are summarized. By matching acronyms with their intended topics using disambiguation algorithms, this problem—also referred to as the Cataphora problem—can be lessened.

Developing strong representations that address issues the system faces and making sure summary phrases have meaning and impact for users are additional challenges. The goal of ongoing text summarizing research is to reach higher degrees of abstraction, which presents linguists and academics with a wealth of options to investigate solutions. [2].

5 Overview of Important Features

Text representation models are becoming more and more common as a means of improving input document comprehension. These models convert words into numerical representations in the field of natural language processing (NLP), enabling computers to identify patterns in language [10]. They create links between certain phrases and the background information in texts. Term Frequency-Inverse Document Frequency, Word Embedding, Bag of Words, and N-grams are a few of the most widely used text representation methods.

1. N-gram: It is perfect for multilingual scenarios since it doesn't require a lot of language preprocessing. To build a model, it divides words or letters into N components. The vector representation that is produced is controllable and of a suitable size. N-grams do have some drawbacks, too; as N increases, their effectiveness rises and they need a significant amount of RAM. Moreover, N-grams produce a sparse language representation based on the probability of term co-occurrence. The chance of words that are not in the training corpus is zero.
2. Bag of words (BoW): Vector sparsity brought on by BoW may have an effect on model performance. Word order inside sentences is also ignored by BoW, despite the fact that it can be important for text comprehension. More sophisticated models like N-grams and Word Embedding have been created to overcome these restrictions.

3. TF-IDF: While TF evaluates a word's occurrence in a particular document, IDF considers a word's frequency throughout the corpus to determine its relevance. TF-IDF calculates a term's importance within a document in relation to its significance throughout the entire corpus by integrating the two metrics. Nevertheless, TF-IDF has limitations, including poor performance in large vocabulary sets and the presumption that term counts serve as separate measures of similarity.
4. Word Embedding: Words or phrases are mapped into fixed-dimensional vectors inside of a continuous space using word embedding methods. Closely spaced vectors depict similar words or phrases [12] and [16]. Few famous algorithms are:

   – FastText
   – Global Vectors for Word Representation (GloVe)
   – Word2Vec

6 Discussion

In this study, we have emphasized the ways in which this technology tackles the problem of distilling large amounts of textual data into clear, short summaries. This has broad uses in business, social media, research, journalism, and other fields.

We have analyzed text summarization from an extractive and abstractive perspective. The main goal of extractive techniques is to take important passages or phrases straight out of the original text and arrange them so that they produce a summary. In contrast, summaries produced by abstractive methods entail interpreting and rephrasing the text in a way that is more human. Emphasis is placed on how abstractive summarization is versatile and flexible.

The article discusses numerous methods and strategies for text summarizing. Techniques including TF-IDF, clustering, graph-based strategies, and deep learning models are mentioned. These methods provide summarizing practitioners with a wide range of tools, accommodating various requirements and levels of complexity in summary jobs.

Finally, we have noted a number of difficulties with text summary. These include the subjective nature of sentence selection, the difficulty of assessing the quality of automatic summaries, and linguistic obstacles like anaphora and cataphora problems. Recommendation [3]. It also covers how to deal with complex language and technical jargon. Our video-based text summary method has a lot of room for growth and innovation in the future. We can provide customers with increasingly sophisticated, adaptable, and context-aware video summarizing tools by consistently pushing the envelope.

Text summarization systems are essential for improving the functionality of many natural language processing (NLP) applications, including machine translation [9], chatbots [11], and annotation systems. These systems enable effective information processing and analysis across several NLP disciplines by extracting and presenting the most important information from massive amounts of text.

The article offers a thorough analysis of text summarization [14], stressing its benefits, different approaches, difficulties, and exciting potential directions for future research. The statement highlights the significance of this technology in meeting the increasing demand for effectively extracting information from extensive textual data sources.

7 Conclusion and Future Scope

Regarding the future, our novel method of video-based text summarizing has intriguing opportunities for further study and advancement. Even though we have come a long way in terms of textual content extraction, audio extraction, and video summarization, there are still a number of exciting directions that could use more research and development, including enhanced multimodal analysis, real-time summarization, customizable summaries, multilingual summarization, deep learning advancements, and application diversification.

1. **Enhanced Multimodal Analysis:** Increasing our capacity to incorporate visual components in addition to audio in videos will offer a deeper comprehension of the information. Richer and more contextually relevant summaries can result from the incorporation of object detection and picture recognition systems.
2. **Real-Time Summarization:** It will be essential to develop real-time video summary algorithms for live broadcasting, social media streaming, and monitoring, among other uses. Processing speed optimization will be necessary for this without sacrificing the quality of the summaries.

3. **Customizable Summaries:** Technology may be made more user-centric by giving users the ability to select the specific features and degree of detail they want in their summaries. Customizing summaries to suit each person's requirements and tastes will be a useful tool.
4. **Multilingual Summarization:** By adding support for more languages, we can make our system more widely useful and accessible by enabling users to summarise videos in languages other than English.
5. **Deep Learning Advancements:** Investigating the integration of the newest models and architectures can result in notable gains in summarization coherence and accuracy as deep learning techniques continue to advance.
6. **Application Diversification:** There will be more chances for business and societal influence if the many uses of video summarization—including content suggestion, education, and market research—are investigated further.
7. **Ethical Consideration:** In order to ensure responsible implementation of video summarizing technologies, it will be imperative to address ethical considerations related to privacy, prejudice, and misinformation detection. It can also be utilized to assess the user's emotional and mental well-being, which will aid in our assessment of the user's psychological well-being.

References

1. Aker, A., Cohn, T., Gaizauskas, R.: Multi-document summarization using a* search and discriminative learning. In: Proceedings of the 2010 conference on empirical methods in natural language processing. pp. 482–491 (2010)
2. AL-Banna, A.A., AL-Mashhadany, A.K.: Natural language processing for automatic text summarization [datasets]-survey. Wasit Journal of Computer and Mathematics Science **1**(4), 156–170 (2022)
3. Bala, A., Mitra, R., Mondal, A.: Recommendation system to predict best academic program. In: 2023 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech). pp. 1–6. IEEE (2023)
4. Dey, M., Mondal, A., Das, D.: Ntcir-12 mobileclick: Sense-based ranking and summarization of english queries. In: Ntcir (2016)
5. Edmundson, H.P.: New methods in automatic extracting. Journal of the ACM (JACM) **16**(2), 264–285 (1969)
6. Ferreira, R., de Souza Cabral, L., Lins, R.D., e Silva, G.P., Freitas, F., Cavalcanti, G.D., Lima, R., Simske, S.J., Favaro, L.: Assessing sentence scoring techniques for extractive text summarization. Expert systems with applications **40**(14), 5755–5764 (2013)
7. García-Hernández, R.A., Ledeneva, Y.: Word sequence models for single text summarization. In: 2009 Second International Conferences on Advances in Computer-Human Interactions. pp. 44–48. IEEE (2009)
8. Kupiec, J., Pedersen, J., Chen, F.: A trainable document summarizer. In: Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 68–73 (1995)
9. Mahata, S.K., Mondal, A., Dey, M., Sarkar, D.: Sentiment analysis using machine translation. In: Applications of Machine intelligence in Engineering, pp. 371–377. CRC Press (2022)
10. Mondal, A., Cambria, E., Dey, M.: An annotation system of a medical corpus using sentiment-based models for summarization applications. In: Computational Intelligence Applications for Text and Sentiment Data Analysis, pp. 163–178. Elsevier (2023)
11. Mondal, A., Dey, M., Das, D., Nagpal, S., Garda, K.: Chatbot: An automated conversation system for the educational domain. In: 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP). pp. 1–5. IEEE (2018)
12. Mondal, A., Dey, M., Mahata, S.K., Sarkar, D.: An automatic summarization system to understand the impact of covid-19 on education. In: Applications of Machine intelligence in Engineering, pp. 379–386. CRC Press (2022)
13. Mridha, M.F., Lima, A.A., Nur, K., Das, S.C., Hasan, M., Kabir, M.M.: A survey of automatic text summarization: Progress, process and challenges. IEEE Access **9**, 156043–156070 (2021)
14. Sinha, S., Mandal, S., Mondal, A.: Question answering system-based chatbot for health care. In: Proceedings of the Global AI Congress 2019. pp. 71–80. Springer (2020)
15. Syed, A.A., Gaol, F.L., Matsuo, T.: A survey of the state-of-the-art models in neural abstractive text summarization. IEEE Access **9**, 13248–13265 (2021)
16. Tas, O., Kiyani, F.: A survey automatic text summarization. PressAcademia Procedia **5**(1), 205–213 (2007)
17. Zhang, Y., Zincir-Heywood, N., Milios, E.: Narrative text classification for automatic key phrase extraction in web document corpora. In: Proceedings of the 7th annual ACM international workshop on Web information and data management. pp. 51–58 (2005)