Automated Text Summarization: A Comprehensive Review

LiLy Turner[1]

[1] Lakeside University-Canada, lilyturner83@outlook.com, https://orcid.org/0009-0003-1119-5245

ABSTRACT

The contemporary era is characterized by an excess of information, wherein data pertaining to a single subject requires considerable time for manual condensation due to its extensive volume. To tackle this challenge, methodologies for Automated Text Summarization have been devised. Presently, two predominant techniques are employed: Extractive and Abstractive methods. This study conducts a comprehensive review of publications within the IEEE and ACM libraries that focus on Automated Text Summarization in the English language.
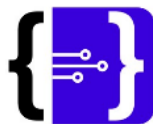
Keywords: *Automated document summarization; extractive approach; abstractive method; literature review*

Resumen Automatizado de Textos: Una Revisión Integral

RESUMEN

La era contemporánea se caracteriza por un exceso de información, donde los datos relacionados con un solo tema requieren un tiempo considerable para su condensación manual debido a su extenso volumen. Para abordar este desafío, se han ideado metodologías para el Resumen Automatizado de Textos. En la actualidad, se emplean dos técnicas predominantes: los métodos extractivos y abstractivos. Este estudio realiza una revisión integral de publicaciones en las bibliotecas del IEEE y ACM que se centran en el Resumen Automatizado de Textos en el idioma inglés.

Palabras clave: Resumen automático de documentos; enfoque extractivo; método abstractivo; revisión de literatura

INTRODUCTION

In the contemporary era, the Internet has evolved into a crucial information source. As we navigate the web seeking information on specific topics, search engines inundate us with an overwhelming volume of data. This surge in information has underscored the significance of Automatic Text Summarization (ATS) as a time-saving tool, offering valuable insights into the essence of information.

The inception of ATS dates back several years, with heightened interest beginning around the year 2000. The early 21st century witnessed the emergence of new technologies in Natural Language Processing (NLP), contributing to the advancement of ATS. Positioned within the realms of NLP and Machine Learning (ML), ATS is defined in the book [11] as the process of distilling crucial information from a source to produce a condensed version tailored for specific users and tasks.

A common approach to summarizing content involves thoroughly reading the text and then expressing the main idea either using the same words or rephrasing sentences. Regardless of the method employed, the primary goal is to capture and convey the most salient idea. ATS aims to create summaries that match the quality of human-generated summaries, with two main methods prevailing: Extractive and Abstractive Text Summarization [3].

Extractive Text Summarization involves extracting keywords, phrases, or sentences from the document, combining them, and incorporating them into the final summary. However, summaries generated through this approach may sometimes contain grammatical errors. In contrast, Abstractive Text Summarization focuses on generating phrases and sentences from scratch, essentially paraphrasing the original document to preserve key concepts in the summary. Notably, summaries produced through this method are grammatically error-free, presenting an advantage over the Extractive approach.

Crucially, summaries generated through Abstractive Text Summarization closely resemble those created by humans, as they involve rewriting the entire document's ideas using a different set of words. It is imperative to acknowledge that implementing Abstractive Text Summarization is more challenging compared to the Extractive approach.
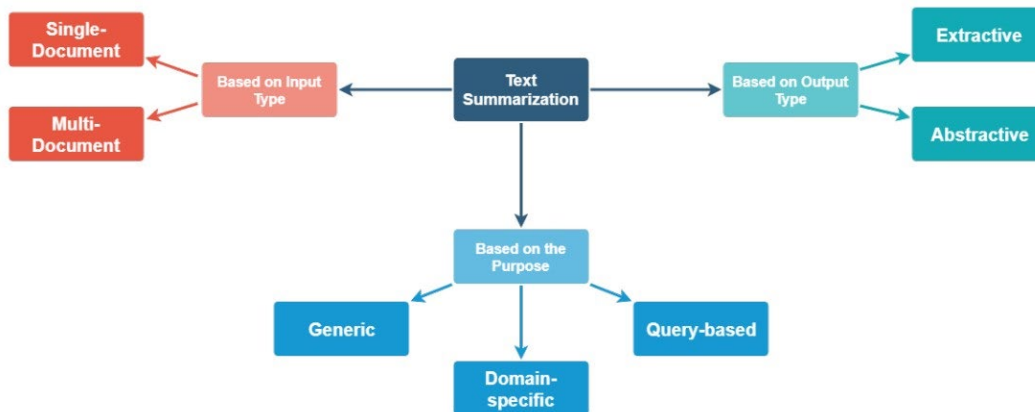
Figure 1 illustrates the various categories of text summarization

Background

The concept of Automatic Text Summarization (ATS) gained traction years ago, but its research momentum increased significantly with the availability of tools for Natural Language Processing (NLP) and Machine Learning (ML), including Text Classification and Question-Answer systems. Some of the documented advantages of ATS [3] include:

- Reducing the time required for reading documents.

- Simplifying the document search and selection process, especially for research papers.

- Generating less biased summaries compared to human-authored ones.

- Providing personalized information in question-answering systems through customized summaries.

ATS finds applications across diverse fields, such as summarizing news articles and streamlining medical records for efficient communication of a patient's history to medical professionals, saving both time and facilitating better understanding of the patient's condition.

The authors of the book [11] enumerate daily applications of ATS, including:

- Headlines from around the world.

- Outlines for students.

- Meeting minutes.

- Previews of movies.

- Synopses for soap opera listings.

- Reviews of books, CDs, movies, etc.

- Digests for TV guides.

- Biographies, resumes, and obituaries.

- Abridgments like Shakespeare for children.

- Bulletins for weather forecasts and stock market reports.

- Sound bites capturing politicians' perspectives on current issues.

- Histories providing chronologies of salient events.

The resurgence of the Internet in 2000 led to a data explosion, prompting the establishment of two evaluation programs (conferences related to Text) by the National Institute of Standards and Technology (NIST) in the USA: Document Understanding Conferences (DUC) [18] and Text Analysis Conference (TAC) [13]. Both conferences contribute valuable data related to text summaries.

Types of Text Summarization

Text Summarization encompasses various classifications. Figure 1 illustrates these diverse categories [3].

Based on Output Type

Text Summarization based on output type can be categorized into two distinct types [3]:

- Extractive Text Summarization.

- Abstractive Text Summarization.

Extractive Text Summarization

Extractive Text Summarization involves the selection of crucial sentences from a given document, which are then incorporated into the summary. The majority of available text summarization tools follow an extractive approach. Some notable online tools include TextSummarization, Resoomer, Text Compactor, and SummarizeBot. The complete list of Automatic Text Summarization (ATS) tools is available at [16].

The process of analyzing the document is straightforward. Initially, the information (text) undergoes a pre-processing step, where all words are standardized to either lower- or upper-case letters, stop words are removed, and the remaining words are converted into their root forms. Subsequently, various

features are extracted, including the length of the sentence, word frequency, the most appearing word in the sentence, and the number of characters in the sentence. Based on these features, a sentence scoring mechanism is employed to arrange sentences in either descending or ascending order. In the final step, sentences with the highest values are selected for inclusion in the summary.

Figure 2 visually depicts the steps involved in the Extractive Summarization process. The authors in [9] have implemented ATS for multi-document scenarios, each containing distinct headings for individual documents.

Types of Text Summarization

Text Summarization encompasses various classifications. Figure 1 illustrates these diverse categories [3].

Based on Output Type

Text Summarization based on output type can be categorized into two distinct types [3]:

- Extractive Text Summarization.

- Abstractive Text Summarization.

Extractive Text Summarization

Extractive Text Summarization involves the selection of crucial sentences from a given document, which are then incorporated into the summary. The majority of available text summarization tools follow an extractive approach. Some notable online tools include TextSummarization, Resoomer, Text Compactor, and SummarizeBot. The complete list of Automatic Text Summarization (ATS) tools is available at [16].

The process of analyzing the document is straightforward. Initially, the information (text) undergoes a pre-processing step, where all words are standardized to either lower- or upper-case letters, stop words are removed, and the remaining words are converted into their root forms. Subsequently, various features are extracted, including the length of the sentence, word frequency, the most appearing word in the sentence, and the number of characters in the sentence. Based on these features, a sentence scoring mechanism is employed to arrange sentences in either descending or ascending order. In the final step, sentences with the highest values are selected for inclusion in the summary.

Figure 2 visually depicts the steps involved in the Extractive Summarization process. The authors in [9] have implemented ATS for multi-document scenarios, each containing distinct headings for individual documents.
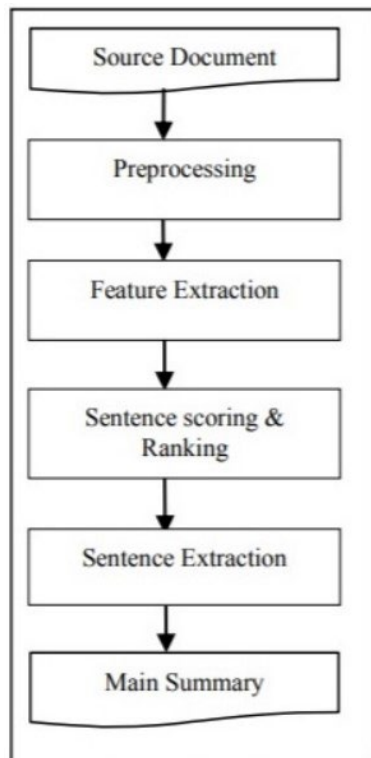

Figure 2 illustrates the stages involved in extractive text summarization.

Abstractive Text Summarization

Abstractive Text Summarization endeavors to replicate a human-like summary by generating novel phrases or sentences, aiming to provide a more cohesive summary to the user. This approach holds appeal because it mirrors the method humans employ to summarize given texts.

However, the challenge lies in the practical implementation of abstractive summarization, which is more intricate compared to the extractive approach. Consequently, the majority of tools and research have predominantly focused on the extractive method. Notably, contemporary researchers are increasingly leveraging deep learning models for abstractive approaches, yielding promising results.

These approaches draw inspiration from the Machine Translation problem. In a notable example, the authors [2] introduced an attentional Recurrent Neural Network (RNN) encoder-decoder model originally used in Machine Translation, demonstrating excellent performance. Subsequently,

researchers adapted the Text Summarization problem into a Sequence-to-Sequence Learning framework.

Another group of researchers [12] employed the same "Encoder-Decoder Sequence-to-Sequence RNN" model, transforming it into an abstractive method for summarization. Figure 3 depicts the architecture of the "Encoder-Decoder Sequence-to-Sequence RNN" model in the context of abstractive text summarization.
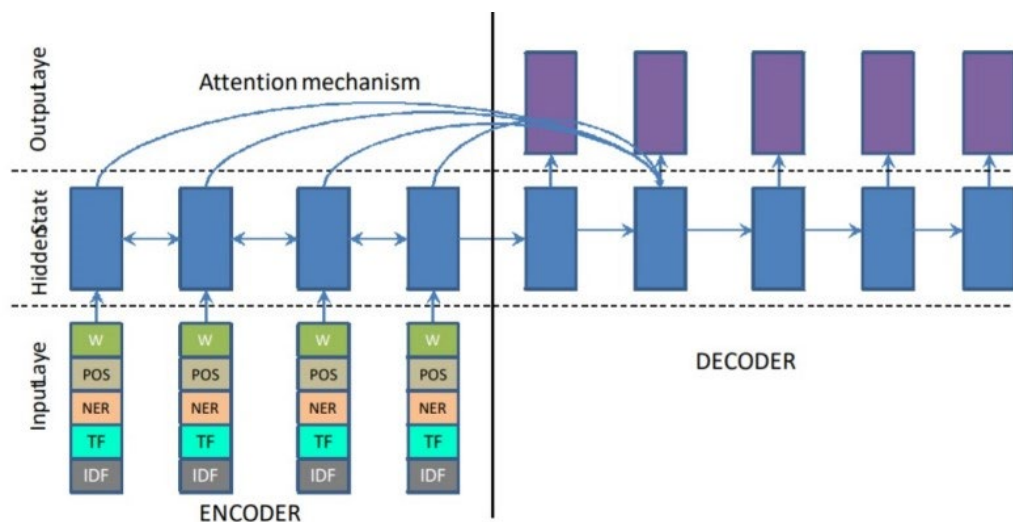

Figure 3 illustrates the structure of the encoder-decoder sequence-to-sequence RNN

Based on Input Type

Text Summarization based on input types can be categorized into two distinct types [3]:

- Single-document.
- Multi-document.

Single-Document

This form of text summarization proves highly effective for condensing information from a single document. It is particularly useful for summarizing short articles or individual PDF or Word documents. Early summarization systems predominantly focused on summarizing single documents. An example of an online tool for single-document summarization is Text Summarizer [17]. This tool accepts a URL or text as input, summarizing the document into concise sentences by identifying and including the most crucial information.

Multi-Document

Multi-document summarization generates a summary from multiple text documents. When provided with various documents related to specific news or articles, this approach creates a succinct overview of the essential events. Multi-document Text Summarization is valuable when users seek to eliminate unnecessary information across various documents, as articles on the same events may contain repeated sentences.

Based on Purpose

Text summarization techniques based on their purpose can be categorized into three distinct types [3]:

- Generic.

- Domain-specific.

- Query-based.

Generic

Generic text summarization is universally applicable, making no assumptions about the domain or content of the text. It treats all inputs equally and is commonly used for tasks such as generating headlines for news articles, summarizing news content, providing overviews of a person's biography, or summarizing sound-bites from politicians, celebrities, entrepreneurs, etc. Much of the work in the field of Automatic Text Summarization (ATS) pertains to generic text summarization. For instance, the authors [15] have developed an ATS specifically designed to summarize news articles, encompassing various categories.

Domain-Specific

In contrast to generic summarization, domain-specific text summarization takes into account the topic or domain of the text. This type of summarization involves incorporating domain-specific knowledge along with the input text to produce a more accurate summary. For example, a text summarization model focused on cardiology might utilize heart-related knowledge, or one in computer science might leverage domain-specific concepts. The key advantage of domain-specific summarization is that the model's understanding of the context is enhanced, allowing it to extract more pertinent sentences related to the specific field.

Query-Based

Query-based text summarization involves users providing a query to the summarization tool, which then retrieves information relevant to that query. This type of tool is primarily employed for natural language question-answering scenarios, aiming to extract a personalized summary based on user needs. For instance, if an article mentions "John" and "Car," and the user wants a summary focusing on "John," the ATS will retrieve a summary tailored to that aspect.

Current Research in Automatic Text Summarization

In recent research on Automatic Text Summarization (ATS), the authors [8] introduced an ATS system designed to summarize Wikipedia articles utilizing an Extractive Approach. Their methodology begins with a preprocessing step, involving text tokenization, the application of Porter stemming, and the extraction of 10 different features (f1 - f10). These features are then fed as input into a neural network comprising one hidden layer and one output layer, producing output scores ranging from 0 to 1. The generated scores are proportional to the sentence importance and are utilized to generate the summary. Additionally, Microsoft Word 2007 is employed to generate a summary for the same article, referred to as the "Reference Summary."

Performance evaluation involves comparing both the Reference Summary and the System-Generated Summary, with precision, recall, and f1-score calculated. The model performs optimally when utilizing only the f9 feature, achieving an f1-score of 0.223, while f7 exhibits the lowest f1-score of 0.055.

In a separate study, other authors [6] presented a 4-dimensional graph model for ATS. Graph models visually represent sentence relationships within a text, offering valuable insights for ATS tasks. The TextRank algorithm is applied in the context of Extractive Text Summarization. Evaluation is performed using the CNN dataset, showcasing overall improvement in the TextRank algorithm—precisely, a 34.87% enhancement in terms of precision, recall, and f-measure compared to the similarity model.

The 4 dimensions utilized to construct the graph are outlined as follows:

- **Similarity:** Measures the overlapping content between pairs of sentences, creating an edge between sentence pairs if it exceeds a user-selected threshold score.

- **Semantic Similarity:** Incorporates ontology conceptual relations, such as synonyms, hyponyms, and hypernyms. Sentences are represented as vectors with words, and semantic similarity scores are calculated using WordNet.

- **Coreference Resolution:** Identifies the noun referring to the same entity through the process of coreference resolution, distinguishing between named, nominal, or pronominal forms.

- **Discourse Relations:** Highlights relevant relationships within the text to provide a comprehensive understanding of the content.

The study by the authors [1] introduces a Query-oriented Automatic Text Summarization (ATS) employing the Sentence Extraction technique. Initial preprocessing steps, including Tokenization, Stop Words Removal, Stemming, and POS tagging, are applied to the input text. Subsequently, 11 features are extracted, with the first set identifying informative sentences and the second set assisting in extracting query-relevant sentences. Each sentence is scored based on these features, utilizing the DUC-2007 dataset for both training and evaluation purposes.

In the work of Min-Yuh Day and Chao-Yu Chen [4], an AI-based approach for ATS is proposed, employing three distinct models: Statistical, Machine Learning, and Deep Learning Models. The dataset comprises essay titles and abstracts. The input, consisting of essay abstracts, is fed into all three models, each generating a headline for the essay. Evaluation involves using the ROUGE evaluation metric, and the best-fitting title is selected from the three generated summaries.

The authors in [5] present an article outlining an unsupervised extractive approach based on graphs. This methodology constructs an undirected weighted graph from the original text by creating a vertex for each sentence and determining weighted edges based on a similarity/dissimilarity criterion. A ranking algorithm is applied, and the most important sentences are identified based on their corresponding rank. The DUC-2002 dataset is utilized for analysis, with results evaluated using ROUGE-1 and various distance measures such as LSA, TextRank, Correlation, Cosine, Euclidean, etc. In a different vein, other authors [7] introduce an ATS grounded in an unsupervised graph-based ranking model. This model builds a graph by capturing words and their lexical relationships from the document, focusing on a subset of high-rank and low-rank words. Sentences are extracted based on the presence

of high-rank words, culminating in the generation of a document summary. Notably, the authors emphasize the application of ATS for individuals with visual challenges or visual loss and evaluate the proposed system on the NIPS (Neural Information Processing System) Dataset, specifically concentrating on Single Document Summarization.
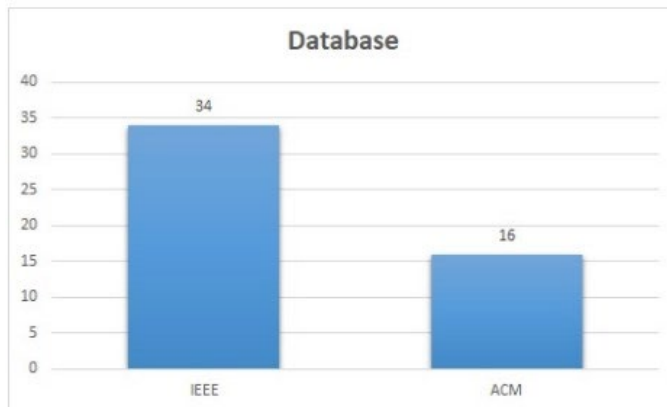

Figure 4 illustrates the distribution of papers across different databases
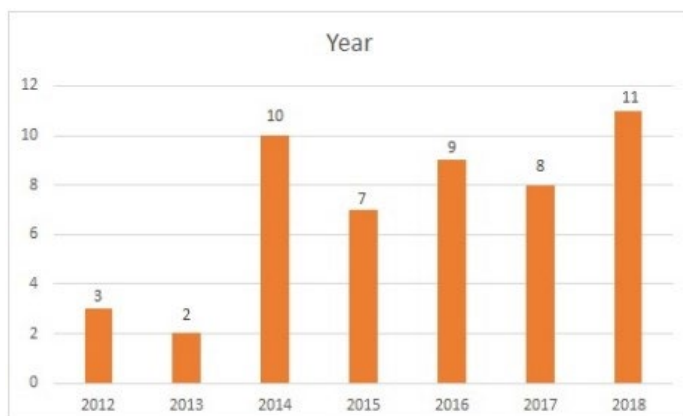

Figure 5 displays the distribution of papers based on the year of publication.

Classification of Papers

The papers selected from two distinct databases were categorized into different classes, and the details are outlined below:

Distribution by Database

A total of 50 research papers were gathered from two databases: IEEE and ACM. Among these, 34 papers (68%) were sourced from IEEE, while 16 papers (32%) originated from ACM (refer to Figure-

4). All the research papers identified were conference papers related to the topic of "Automatic Text Summarization."

Distribution by Publication Year

While querying our topic, over 160 papers were identified, but the search was confined to the past seven years. Figure-5 illustrates the distribution over the years, revealing a relatively low volume of research in 2012-2013, followed by a noticeable increase from 2014 onwards. The data indicates two peak years for publications, with 11 papers in 2018 and 10 papers in 2014.

Distribution by Type of ATS

Based on Output Type

As previously mentioned, there are two types of ATS based on output type: Extractive Text Summarization and Abstractive Text Summarization. Additionally, a "Hybrid" category was introduced for papers employing a combination of Text Summarization techniques. Classification based on these types reveals that among the papers, 46 fall under the Extractive type, 2 under Abstractive, and 2 under Hybrid (see Figure 6).

Based on Input Type

Given that there are two types of Text Summarization based on input type—Single and Multi-Document Summarizations—Figure 7 indicates that 46 papers focused on Single Document ATS, while 4 were centered on Multi-Document ATS. The majority of researchers concentrated on summarizing a single input document.

Dataset

Within the array of papers, the DUC [18] dataset emerged as the most widely adopted among research studies. DUC provides single-document articles paired with manually crafted summaries, often referred to as the "Gold Summary." This gold standard serves as a benchmark for evaluating the effectiveness of the summaries generated through Text Summarization techniques. DUC encompasses various datasets organized chronologically, spanning from 2001 to 2007.

The second most prevalent dataset is the CNN dataset, comprising news and articles sourced from the CNN website. Additionally, several other datasets were utilized in various papers, including but not limited to Gigaword, Elsevier Articles, Opinonis, and Daily Mail.
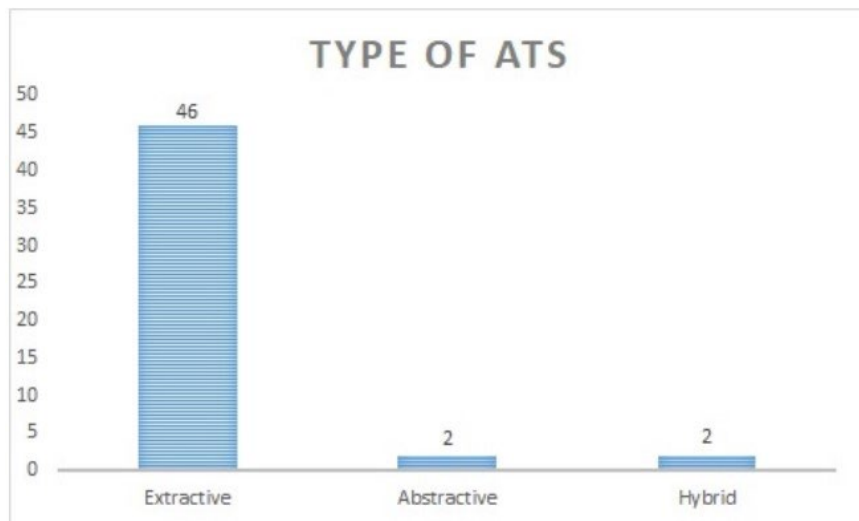

Figure 6 illustrates the breakdown of papers based on the output type of text summarization
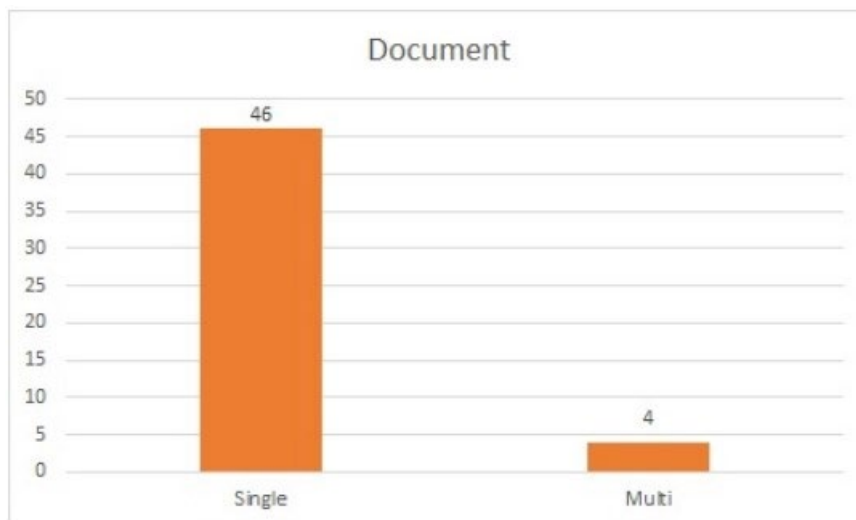

Figure 7 presents the distribution of papers categorized by the input type of text summarization.

Evaluation Methodology

The primary evaluation method employed is the ROUGE evaluation [10], an acronym for Recall-Oriented Understudy. ROUGE is a comprehensive set of metrics designed to assess automatic text summarization in textual content and machine translation.

In essence, the evaluation involves comparing two distinct types of summaries: the Automatically Produced Summary generated by ATS and the Set of Reference Summary, typically crafted by humans. Another evaluation approach utilized is the F-1 measure, wherein Precision and Recall are computed.

The F1 score, or F-measure, serves as a metric to determine the accuracy of a given system. Its computation incorporates both Precision (p) and Recall (r), with Precision representing the fraction of the summary that is correct, and Recall indicating the fraction of the correct (model) summary that is outputted. It's worth noting that certain papers opted not to employ any evaluation metrics for accuracy assessment.

In the context of ROUGE, Precision assesses the extent to which the ATS summary is truly relevant or necessary. On the other hand, Recall evaluates how much of the reference summary the ATS summary is able to retrieve or capture.

$$\text{Precision} = \frac{number\_of\_overlapping\_words}{total\_words\_in\_system\_summary}, \quad (1)$$

$$\text{Recall} = \frac{number\_of\_overlapping\_words}{total\_words\_in\_reference\_summary}. \quad (2)$$

Apart from Precision and Recall, three additional evaluation metrics include:

Rouge-N

This ROUGE package [10] evaluates unigrams, bi-grams, trigrams, and higher-order n-grams overlap. For instance, ROUGE-1 assesses unigrams, while ROUGE-2 focuses on bigrams, and ROUGE-3 targets trigrams. Both the system summary and reference summary are utilized to calculate the overlap of these n-grams [14].

Rouge-L

This metric [10] measures the longest matching sequence of words using the Longest Common Subsequence (LCS). The advantage of LCS lies in its ability to capture non-consecutive matches while maintaining the sequence that reflects sentence-level word order. It inherently includes the longest in-sequence common n-grams, eliminating the need for defining a predefined n-gram length [14].

Rouge-S

Skip-gram assesses the overlap of word pairs with a maximum of n gaps between words. For example, skip-bigram evaluates the overlap of word pairs with a maximum of two gaps between words [14, 10].

CONCLUSION

This study delved into an examination of 50 papers sourced from the IEEE and ACM databases in the realm of Automatic Text Summarization. We provided an overview of diverse ATS types categorized by input, output, and purpose. Additionally, the paper discussed ongoing research studies and categorized the collected papers based on criteria such as year, input type, output type, and the database. The exploration also included a discussion on various databases commonly utilized by researchers. Finally, the paper elucidated the most frequently used evaluation metric, ROUGE, along with its distinct metrics.

BIBLIOGRAPHICAL REFERENCES

1. Afsharizadeh, M., Ebrahimpour-Komleh, H., Bagheri, A. (2018). Query-oriented text summarization using sentence extraction technique. 2018 4th International Conference on Web Research (ICWR), IEEE, pp. 128–132.

2. Bahdanau, D., Cho, K., Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate.

3. Chauhan, K. (2018). Unsupervised text summarization using sentence embeddings.

4. Day, M. Y., Chen, C. Y. (2018). Artificial intelligence for automatic text summarization. 2018 IEEE International Conference on Information Reuse and Integration (IRI), pp. 478–484.

5. de la Pena Sarrac˜en, G. L., Rosso, P.´ (2018). Automatic text summarization based on betweenness centrality. Proceedings of the 5th Spanish Conference on Information Retrieval, pp. 11.

6. Ferreira, R., Freitas, F., Cabral, L. d. S., Lins, R. D., Lima, R., Franc¸a, G., Simskez, S. J., Favaro, L. (2013). A four-dimension graph model for automatic text summarization. 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), volume 1, pp. 389–396.

7. Hamid, F., Tarau, P. (2014). Text summarization as an assistive technology. Proceedings of the 7th International Conference on Pervasive Technologies Related to Assistive Environments, Association for Computing Machinery, pp. 60.

8. Hingu, D., Shah, D., Udmale, S. (2015). Automatic text summarization of Wikipedia articles view document. 2015 International Conference on Communication, Information & Computing Technology, IEEE, pp. 1–4.

9. Krishnaveni, P., Balasundaram, S. R. (2017). Automatic text summarization by local scoring and ranking for improving coherence. 2017 International Conference on Computing Methodologies and Communication, pp. 59–64.

10. Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. Text Summarization Branches Out, Association for Computational Linguistics, pp. 74–81.

11. Mani, I., Maybury, M. T. (1999). Advances in automatic text summarization. MIT press.

12. Nallapati, R., Zhou, B., dos Santos, C. N., Gulcehre, C., Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence RNNs and beyond.

13. NIST (2018). Text Analysis Conference.

14. RxNLP (2018). ROUGE Evaluation Metrics.

15. Sethi, P., Sonawane, S., Khanwalker, S., Keskar, R. B. (2017). Automatic text summarization of news articles. 2017 International Conference on Big Data, IoT and Data Science (BID), pp. 23–29.

16. Softsonic (2018). List of Online Automatic Test Summarization Tools.

17. Text Summarizer (2018). Manual of Text Summarization.

18. Voorhees, E. (2002). Document understanding conferences website.