iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XI **Month of publication:** November 2024

DOI: https://doi.org/10.22214/ijraset.2024.65100

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

An Exploration of Extractive method and Abstractive Method of Text Summarization with Various Approaches, Techniques and Datasets

Prafull S. Mankar [1], Avinash B.Manwar[2]
[1]Research Scholar, [2]Associate Professor, Department of Computer Science, S.G.B.Amravati University, Amravati, Maharashtra, India

Abstract: ***In today's era, an enormous amount of data available which is in complex form such as Social media content, Images, Audio, Video and Text data form. The mechanism is very much needed to provide these types of data in simple and easy to understandable form. In this paper, importance is given on text summarization of large amount of textual data of documents must be specified in concise and understandable form. Automatic Text Summarization is the technique which is used to represent most significant concepts in precise and comprehensible form from original or source document to target document providing greater flexibility and convenience. Since the advancement in emerging technologies leads to the challenges in research field which is leading towards to need arises to analyze and should be resolved over the time period.***
Generally, there are two approaches for automatic text summarization: Extractive and Abstractive. The process of extractive based text summarization can be divided into two phases: pre-processing and processing. Automatic text summarization process model can be divided into three steps in the preprocessing step, source text interpretation to representation, source text representation transform to summary text representation with an algorithm. Finally, summary text generation from summary representation. In this paper discussion were made about some of the extractive based and abstractive based methods of text summarization for single documents. Various approaches of text summarization, methods needed for it along with tools and techniques, metrics needed to measure quality of generated text summary applications have been discussed in detail.
This survey paper highlights the main advancement in the field of Automatic Text Summarization and provides the key methods with their benefits and limitations. Gives an overview of the past, present and future direction with certain solution to the challenges arises in text summary generation process with extractive and abstractive methods.
Keywords: Automatic Text Summarization, Extractive methods, Abstractive methods, Semantic- Based , Topic-Based, Graph-Based , Machine Learning, Recurrent Neural Network

I. INTRODUCTION

The text summarization is the process in which essential information is produce automatically in concise and coherent ways by using different approaches of text summarization. In the text summarization process extractive type of summarization approach creates summaries by selecting a important sentences from the original text. As against to that, in abstractive summarization as we take main ideas, meaning and form concise and coherent summary from original text. But it is having more tricky challenges like lexical representation, deduction, and natural language production [1]. Extractive summaries comprise of sentences taken directly out of the text, whereas abstractive summaries use terms and phrases that are not included in the original source text. Apart from this there is classification of summary so called as generic or query-oriented, where generic summaries offer an overview of the content of the document, while query-oriented summaries provide the most relevant information to answer particular queries[2]. In Natural language processing the text summarization is applicable in many different contexts. A summary conserve the essential details of the original text while providing a simplified version of it.
There are several methods for extractive summarization like word frequency, cue words or phrases, lexical chains, machine learning, sentence compression using syntactical or statistical constraints, and the use of psychological models. Nonetheless, a number of systems have been demonstrated to provide summaries with sufficient amounts of information, especially when included with other systems [3]. A summary can be defined as a text produced from one or more texts that convey important information in the original text, usually no longer than half of the original text and usually significantly less than that.

©IJRASET: All Rights are Reserved

The main goal of a summary is to present the main ideas in a document in less space. If all sentences in a text document were of equal significance, producing a summary would not be very effective, as any reduction in the size of a document would.

The technology of automatic document summarization is capable of providing a solution to the information excess problem by providing a concise summary of each document [4]. The Summary can be classified into descriptive, evaluative, indicative, and informative types of summary. In Descriptive summary, it describes both the form and content of a source text, while evaluative summary offer critical responses to the source. On the other hand, Indicative summary provides abbreviated information on the main topics of a document, preserving its most significant passages and often used as the end part of information retrieval systems. An Informative summary provide a reduction for full documents, retaining vital details while decreasing information quantity. Mainly there are two approaches for document summarization i.e supervised and unsupervised. Supervised approaches indulges document summarization as a task that requires a high level of knowledge and can be applied to various types of information retrieval systems [5].

A vast amount of electronic documents are available online, making it tricky for users to pull out useful information. This has led to a lack of relevant and interesting documents. As the Internet grows exponentially, it becomes difficult to identify relevant information to satisfy users' information requirements. Automatic document summarization, united with conventional information search engines, can help reduce information overload and efficiently access the relevance of retrieved documents. Tools that automatically produce summaries are essential for professionals and bulky searching engines like Google, AltaVista etc.[6] These tools can be single-document or multi-document, depending on the number of documents to be summarized. Multi-document summarization is an extension of single-document summarization, used for precisely describing the information contained in a cluster of documents and helping users understand the document cluster. It performs knowledge synthesis and knowledge finding, and can be used for knowledge acquisition **.**

II. RELATED WORK

Over the years, Text Summarization approaches have changed, with new machine learning algorithms being proposed for tackling text summarization. The Multi-document summarization and multi-lingual summaries have gained importance due to the vast amount of information over the Internet. There is a trend towards sentiment-based or personalized summaries, and abstract generation is becoming important [7]. The most important text summarization strategies from the last decade, assessing techniques on three different criteria is to chooses four methods mainly Word Frequency, TF/IDF, Lexical Similarity, and Sentence Length [8]. The abstractive based summarization neural attention based model combine with neural machine translation with generation algorithm with plans to improve grammatically produce grammatical level summaries [9].It has been observed that an extractive based linguistic preprocessing steps, tools and plugins providing details about English text features with lexical chain approach to evaluates system and human generated summary [10]. Leading to that some Indian and foreign language with requirement of single summarizer used and tested for various content types especially for Hindi language sentences comparing machine learning techniques on dataset like news and autobiography [11].

It has been the important aspect is to focus on work in keyword extraction, text databases, summarization processes, methodologies, and evaluation matrices etc. Thus highlights text summarization in low resourced languages like Telugu, Hindi, Tamil, and Bengali, multi-lingual text summarization, multimedia summarization, and multi-lingual multimedia summarization [12]. Some of the methods like Fuzzy login have been use for extractive summarization in more qualitative way. The extraction based summarization identifies relevant sentences with Fuzzy logic as a sub task which enhances text summary quality. The online summary with Latent system analyzed semantic relation between ideas in text use by sentence function extraction Fuzzy logic system [13]. The summary evaluation methods, both intrinsic and extrinsic, and the evaluation programs that have occurred which provides a list of advantages and disadvantages of each technique with description of abstractive and multilingual techniques, comparisons of techniques. The best score for summary generation through clustering method and progressive and optimization based approaches has been resulted from the evaluation [14].After this the states of the art in graph summarization provided by dividing input graphs into various types and summarization techniques. Also algorithm, methods and evaluation techniques with key details with comparison is provided such as graph types, such as multilayer and multi-view graphs etc. It also highlights graph summarization using deep node representation despite of growing volume of data [15]. It can use large scale dataset for extreme summarization pushing boundaries of abstractive methods and resulted in better performance recognizing relevant content along with informative summaries [16]. To overcome unique limitations in document processing, algorithms, and outputs some suggestion like clustering with cosine similarity algorithm for sentence extraction, NEWSUM algorithm for generating clusters and position score algorithm for ranking extracts sentences used for improving efficiency and effectiveness of summarization system [17].

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

In multi-sentence text abstractive summarization the effectiveness of temporal hierarchical network uses multiple timescales with adaptation. For long multi sentence text abstractive text summarization performance is improved by using temporal hierarchical networks [18].

An unsupervised framework such as SummCoder for extractive summarization using deep auto encoders and sentence embedding for single document extracting by ranking and sentence selection from input text. There are sentences content relevance, sentence novelty and sentence position combine to generate final sentence selection score. The summary generated by ranking sentence with weighted fusion of scores with this matrices. The overall approach is unsupervised and does not require gold summaries to train the text summarization system [19]. The text summarization using ontology and page rank approach and WMD word machine word embedding used to take sentence decision matching in a represented graph task. For evaluating valuable sentences PageRank algorithm was used along with extractive ontology graph-based text summarization model, SumOnGraph [20]. Using LSTM encoding and decoding for English to English text summarization provides fluent summary with minimal training loss. The main aim is to develop to learn and generate summary with machine learning and deep learnin [21].The utilization of machine learning and deep learning in generating summary is important. The two stage sequence to sequence model using BERT is evaluated on CNN/Daily mail and New York Times datasets and model applied to generate task like machine translation, question generation and paraphrasing for NLP [22]. The event based ATS approach has been proposed to enhance abstract of scientific papers literature and consideration were made about classification into problem, method, result and others. TextRanks algorithm combine clue word and sentence length to compute sentence importance and specify limitations include small corpora size for constructing trigger word templates, small test dataset size, and domain-specific trigger word templates [23].

A Hybrid approach for query relevant text summarization with LSA techniques enhances and reduces redundant data without supervised learning by using PageRank algorithm. However, using only semantic and graph-based methods doesn't solve query-based text summarization completely [24]. A Novel model for Single document summarization uses extractive ATS methods like graph-based, statistical-based, semantic-based, and centrality-based is evolved. The framework also includes a text graph model which dynamically computed weight values for graph nodes, and algorithms for constructing the text graph model and searching for candidate edges and getting candidate summary. EdgeSumm has been evaluated using the ROUGE tool and has shown promising results in single-document generic summarization [25]. Text summarization is essential for saving users time and resources because of abundant data availability. Various algorithms and methods used to produce various types of summaries, with accuracy scores like ROGUE and TF-IDF for better and more concise summaries. These summaries may not always relevant to the original document and there is no specific model for generating the best summaries, but involve modifying models like GANs and transfer learning for more accurate summaries [26]. To provides abstractive text summarization with analysis by mechanism like encoder-decoder along with training methods, dataset and metrics for evaluation. Also gaps review like use of pre-trained models like BART and MASS explores attention mechanism for sequence to sequence model [27]. For enhancing process of text summarization, K-means clustering algorithm combines Gensim Word2Vec and evolve a new sentence scoring procedure. By clustering algorithm model cluster the sentences based on numerical values and nouns. For that BBC news article datasets has been used for Extractive multi document summarization [28]. Again for academic articles and financial reports long documents summarization DANCER a novel summarization method is used. For achieving better performance Seq2Seq RNN model combine with PEGASUS so as to improve Rouge scores [29]. Tree-based Key Extraction technique (TeKET) uses statistical knowledge without requiring train data by introducing variant of binary tree, KeyPhrase Extraction from candidate KeyPhrases. The technique outperforms in terms of F1-scores and domain and language independence [30]. Unlike graph based method, the techniques only compares each sentence with topic vector once so that time saved compare to learning RNN,CNN and LSTM. The method is flexible and adaptable for future changes. The method has been implemented and evaluated for the natural English language and can be adapted for any other language [31].

The query based, a word sense disambiguation method for extractive text summarization introduce to help for finding sense - oriented query relevance based on their meaning. The proposed method finds query based word's sense with semantic relatedness scores between words not presents in WordNet Ontology [32].For diversity, redundancy and compression rate to decide the meaningfulness of a summary, the method introduce document embedding based text summarization approach for capturing semantic relatedness between sentences to generate quality summary. The model is tested on two datasets in English and Hindi, resulting in generating a summary rich in diversity, compressed, and ignores unnecessary text. The model is scalable for more languages and is reliable [33]. The focus is on deep neural sequence-to-sequence models, reinforcement learning approaches, and transfer learning approaches of abstractive ATS. [34]. Extractive text summarization in Natural Language Processing (NLP), focuses on providing non-redundant, short, logical, and useful information from documents in the form of summaries.

The quality of summary is achieved by using evaluation measures and by the techniques from range from basic statistical-based to complex NN and optimization-based approaches [35]

III. APPROACHES FOR TEXT SUMMARIZATION

There are five types of extractive approaches suitable for summarization-

1) *Statistical based Approach:* In this approach, term frequency count engaged to determine the significance of a sentence in scientific documents. However, these methods may not be adequate for building high-quality summaries. Mutual information measures common information between two words, while information gain is a high-quality metric for deciding the relevance of an attribute. Residual inverse document frequency computes the term document frequency according to the Poisson distribution. To determine the relevance each technique assigns weights to the words included in the document, and sentences are scored based on these weights.
2) *Topic based approach:* The topic based approach have five different ways of representation such as topic signatures, enhanced topic signatures, thematic signatures, modeling the content structure of documents, and using templates. In addition to topic words, other approaches consider informativeness and event words to produce multi-document summaries.
3) *Graph-based Approach:* Graph-based ranking algorithms have been shown to be effective in text summarization (TS). The graph node represents text elements, while graph edges are links between them. LexRank is a multi-document summarization system that connects sentences if their similarity is above a predefined threshold. Analyzed several graph-based algorithms and their application to automatic sentence extraction in text analysis.
4) *Discourse-based approaches:* In this approach, the summarization can be employed to address the problem from a linguistic viewpoint. Rhetorical Structure Theory (RST) is a basis for the summarization approach developed, which uses discourse representation which determines the most important textual units in a document. RST with a generic summarizer to add linguistic knowledge to the summarization process and approach similar to RST, focusing on the coherence and cohesion of a document [7][36]. A combination of graph-based with concept-based approaches for document summarization also used. But Graph-based models are extensively used in document summarization because of their ability to efficiently represent document structure. Another graph-based approach is LexRank based on Eigen vector centrality which determines sentence salience. Sentences are clustered into groups based on their similarity measures and ranked based on LexRank scores.

- Fuzzy logic-based approaches consist of four components: defuzzifier, fuzzifier, fuzzy knowledge base, and inference engine. These methods use textual characteristics input, such as sentence length and sentence similarity for improving the quality of summarization. A fuzzy logic approach for ATS, involves pre-processing the text document, feature extraction, and ranking sentences in order to maintain coherency.
- Concept-based approaches extract the concepts from external knowledge bases like HowNet and Wikipedia. These methods estimate sentence significance based on the concepts retrieved from HowNet, resulting in a conceptual vector model for rough summarization. However, these approaches have restrictions, such as a large set of training data, human interruption needed for sentence selection, slow training, and difficulty in adjusting to user requirements.
- Latent Semantic Analysis (LSA): It is an unsupervised learning approach which specifies the hidden semantic structures in sentences and words commonly utilized in text summarization. It captures the text of the input document and extracts information such as frequent words and words seen in different sentences. To find interrelations between words and sentences, for improving accuracy LSA uses Singular Value Decomposition (SVD).

5) *Machine learning approach:* It is based on Bayes rule uses a set of training documents and extractive summaries as input. The sentences are classified as non-summary and summary sentences based on the features of the sentence.

- A neural network-based approach: It uses a two-layer neural network with back propagation trained using the RankNet algorithm. First the labeling training data done and extract features in both test as well as train sets by using machine learning approach. Another method is to use a three layered feed forward neural network which classify summary and non summary sentences. Conditional random fields are a statistical modeling approach which focuses on machine learning to endow with structured prediction. This method identifies correct features and provides a better representation of sentences and groups into segments. However, it is domain-specific and need external domain-specific data.[37]
- Heuristic-based approaches: In the Research on Automatic Keyword Extraction (AKE) employs heuristic-based approaches like keyphrase identification using human annotated samples, threshold-based approaches, pattern filtration, and a four-step process called TAKE. Supervised approaches, such as KEA, KEA++, and feature-based approaches used to improve performance.

KEA utilized word features such as TF-IDF and occurrence information, and implements a Naive-Bayes classifier for keyphrase extraction.

- Neural network approaches: Neural network approaches, such as Multi-Layer Perceptron (MLP) and Recurrent Neural Network (RNN), have also been used in AKE research. A multi-layer perceptron system compares its performance with a simple TF-IDF-based approach. It used a multi-layer perceptron system, comparing its precision and recall with author-generated ones. Compared a range of ML-constructs, including Naive-Bayes, Decision Trees, and Artificial Neural Networks, and establish that the MLP approach outperformed the other two classifiers [38].
- Semantic approach: This approach is crucial especially for multi-document summarization; the context of meaning and redundancy in text is realized in it. Semantic role labeling (SRL) is used to comprehend sentences in a document, creating clusters of identical elements to form summary based on selected sentences. The objective is to modify the text with a predicate argument structure, annotating the verb and identifying its action as a propositional argument called a frame. LSA is an unsupervised approach for extracting semantics of text based on noticed words. Abstractive summarization requires semantic understanding of the document, and authors have developed a semantic graph using SRL [39].
- Numerical based approaches: Numerical based approaches allocate numerical scores to text elements, while relevance of sentences condenses a text based on the impression of MMR between words or sentences. Graph theoretic approaches extract semantically alike sentences and establish connections between them using edges between nodes. Sentence clustering approaches use WordNet and stop word lists to yield a summary from the input text.

A hybrid approaches numerical with linguistic, proposes mining meaningful data from a text document and summarizing it automatically using swarm optimization algorithms. Automatic Text Summarization (ATS), which has three main stages: pre-processing, adaptation of PSO algorithm, and PSO-based data clustering [40]. Extractive summarization uses statistical and heuristic methods to identify and extract important text from the original document, without changing it. It is easy to achieve but faces issues like ambiguity and miscommunication. Abstractive summarization first understands the document and then generates a summary by introducing new words, sentences, and rephrasing. It generate a more relevant and accurate summary with reduced ambiguity and a good compression rate [41].

Table 1. Automatic Text Summarization Approaches Summary

Approaches	Description
Graph-based Approaches	This approach utilizes graph theory to represent semantic relationships between document elements. It Represents textual units and relationships as undirected graphs. It considers cosine similarity and frequency of common terms. It improves coherency by identifying redundant information. It then computes similarity score without considering word significance.
Semantic-based Approaches	Semantic based approach uses TextRank for ATS and keyword extraction process. In this each sentence is represented by a node in the graph and the edge connecting them indicates the similarity relation. LexRank uses a cosine similarity measure of TF-IDF vectors and TextRank uses its similarity measure on the quantity of words shared between two phrases.
Multimedia Semantic Graph	This approach utilizes for display of web document and the summary generation. Through statistical and semantic analysis of textual and visual components it creates a Visual Semantic Tag Cloud. It enhances user knowledge attainment through a synthesized visualization by using semantic analysis.
EdgeSumm	Uses NLP techniques like lemmatization and POS tagging to create a new graph representation of the input text. It Identifies potential sentences for the summary by combining the sentences and rearranging them. EdgeSum works with any domain that document addresses.
Semantic-Based	Semantic-based techniques represent documents to be summarized using a semantic

Approaches to Summarization	representation. LSA creates a semantic representation based on observed co-occurrence of terms. Semantic Role Labeling (SRL) is used for sentence semantic parsing. The summary's quality is determined by quality of the input document's semantic representation. WordNet-based semantic approaches are presented for ATS.
Topic-based Approaches	Topic Aspect-Oriented Summarization (TAOS) was created to characterize discrete themes. The system extracts several feature groups and chooses a common attribute group based on latent variables and a chosen group norm punishment. A greedy technique was used to construct the summary while taking diversity and coverage into account.
Machine Learning Approaches to Summarization	Machine learning methods represent summarization as a supervised classification problem and train it using a training set of data. The system characterizes between summary and non-summary sentences based on humanly created summaries. Machine learning-based systems require a large quantity of training data, making sentence selection for the final summary demanding. The extractive ATS approach uses a novel text representation model and a classification method called the cascade Multimedia Tools and Applications forest. Numerous machine learning approaches for ATS including the Gaussian mixture model, feed forward neural network, probabilistic neural network, genetic algorithm, and mathematical regression is used.
Deep Learning Approaches	Deep neural network-based method computes a feature space from the input representation based on a document phrase using an unsupervised method based on deep auto encoder (AE).The ranking of sentences is based on how closely their meaning matches the query. An Ensemble Noisy Auto-Encoder (ENAE) utilized to enhance the method.

Recent Automatic Text Summarization Approaches [42]

IV. ATS TOOLS AND LIBRARIES

1) *Sumy:* The Sumy is a Python library for extracting summaries from HTML web pages or plain texts. It offers various algorithms and methods for text summarization, including Luhn heuristic method, Latent Semantic Analysis, Edmundson heuristic method, LexRank, and TextRank. Using LexRank: An unsupervised approach inspired by PageRank and HITS. The sentences using the graph method, calculating importance based on eigenvector centrality. Then Sentences recommends other similar sentences to the reader. SPACY: It is a free, open-source library for advanced Natural Language Processing (NLP) in Python. Specially designed for production use and helps to build applications that process large volumes of text. It also supplies linguistic annotations to provide insights into a text's grammatical structure.
2) *GENSIM:* It is fastest library for training vector embeddings in Python or any language. Runs on all platforms including Linux, Windows, OS X, and any other platform that supports Python and NumPy.**NLTK:** A summarizer with the help of the NLTK library, Tokenizes the corpus and removes stop words [43].

V. TEXT SUMMARIZATION APPROACHES /METHODS/TECHNIQUES

Comparative Study of Abstractive Text Summarization Methods

Methods	Description
Structure-Based Approach	It classifies reports based on classifications. It outlines questions based on classifications. It can create rules based on these questions. Also uses set choice module to select the best opponent. It uses long-term configuration plans for outline sentences.
Ontology Method	It reflects neighborhood cosmology of information times in nearby specialists. It uses extraction to extract cooperation degree.

Tree-Based Method	This method is used to preprocesses the same sentences using shallow analysis. It uses algorithms to select a common phrase in sentences. It is useful for FUF / SURGE language generator to create new abstract sentences. It also increases language fluency and reduces system errors.
Multimodal Semantic Models	It can capture ideas and arranges interactions between these ideas. It approves archives and photo records. It has three stages mainly semantic model ideas, rated concepts and sentence generation. It uses information item-based method instead of hypothetical from data archive sentences.
Template-Based Methods	It generates cohesive summaries and explanatory summaries. In this the predefined templates limit variation.
NER Summarisation	SpaCy library is fastest and suitable for practical applications and Flair library outperforms.It is proficient for experiments. In this the challenges include language uncertainty and quality and fidelity of annotation.
Sequence to Sequence (RNN) Summarisation	This is suitable for short sentences but requires large structured data for training. Also the RNN-based Seq2Seq models take long to train and cannot capture distant dependence links for lengthy sequences.
Semantic-Based Methods	It determines semantic link between sentences words.

Table 2. Comparative Study of Abstractive Text Summarization Methods [44][45]

VI. STANDARD DATASETS IN ABSTRACTIVE TEXT SUMMARIZATION

1) *DUC/TAC:* This is extensively used datasets in the Automatic Text Summarization research field which releases by the National Institute of Standards and Technology (NIST).
   In this, each item contain a document and its equivalent ground-truth summary.
   It can also be used as the testing set to estimate the performance of the Automatic Text Summarization model.
2) *CNN/Daily Mail:* It is used in passage-based question answering systems.In the field of abstractive text summarization it can be most widely used benchmark and dataset. It contains two most popular versions: Anonymized Version and Non-anonymized Version.
3) *Gigaword:* It can be comprises of nearly 10 million English news documents from different news agencies used for training and testing neural network models. Gigaword known for the headline (title) generation task reason.
4) *NYT:* It can be comprise of millions of articles in the New York Times between 1987 and 2007 used for automatic summarization, text classification, content extraction, and other NLP tasks. It is preprocessed to make it suitable for text summarization process or task.NYT dataset summaries are more diverse in nature , shorter in length, and make the most of higher levels of abstraction and paraphrase[46].

VII. CONCLUSION AND FUTURE DIRECTIONS

Automatic Text Summarization system is very crucial for accessing information due to vast amount of information overload over the networks. Automated summary generation is a problem in NLP. Automatic text summarization is valuable for Natural Language Processing tasks like query answer and text categorization, as well as connected fields like information retrieval and exploration time improvement. This comprehensive survey paper provides an overview of Automatic Text summarization along with types, classifications, approaches, applications, methods, implementations, datasets, and evaluation measures. A thorough understanding of context and semantics is necessary to ensure accuracy while producing human-like summaries using extractive and abstractive summarization.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The paper also discusses a type of information that needs to be extracted and processed differently from factual data available on the internet. Another problem is dealing with lexical and contextual ambiguities in language. The future of ATS is defined by both massive promise and problems, with large-scale pre-trained language models. The coherence aspect of the summary is best highlighted using sentimental analysis. This paper presents an extractive and abstractive approach to text summarization, analyzing and comparing supervised and unsupervised learning algorithms.

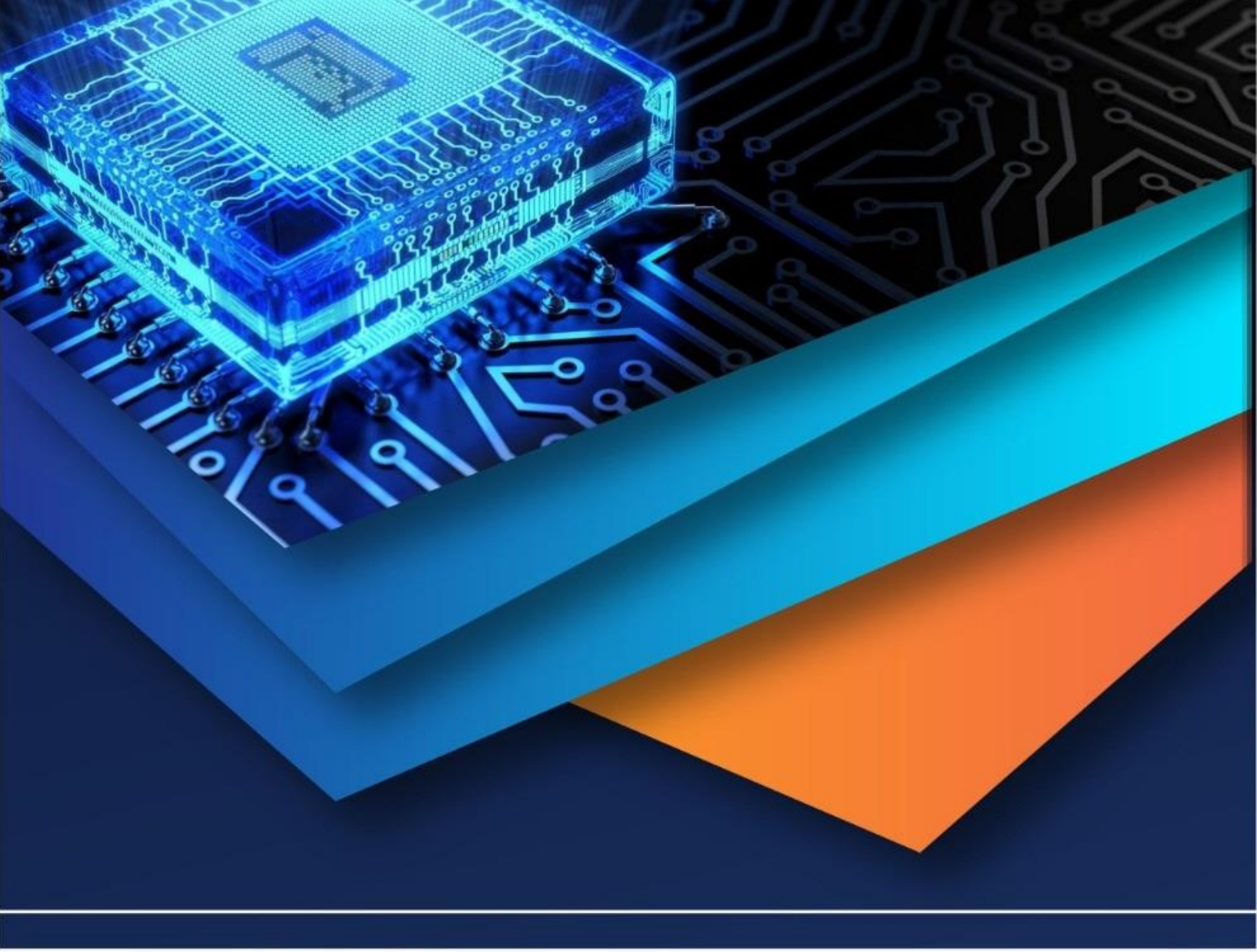
Text-based methods like Neural Network, Graph Theoretic, Fuzzy, and Cluster have been effective, but inconceivable summaries require heavy equipment and are complicated to replicate in certain domains. The future of text summarization shows potential, as researchers persist to explore new techniques and methods to improve the accuracy and efficiency of summarization systems by using deep learning-based summarization. Further research is needed to address these challenges and develop more effective ATS systems for various applications.

REFERENCES

[1] G. Erkan and D. R. Radev, "LexRank: Graph-based Lexical Centrality as Salience in Text Summarization," jair, vol. 22, pp. 457–479, Dec. 2004, doi: 10.1613/jair.1523.
[2] Dou Shen, Jian-Tao Sun, Hua Li, Qiang Yang, Zheng Chen, "Document Summarization using Conditional Random Fields", IJCAI-07, pp. 2862- 2867, 2007.
[3] L. Antiqueira, O. N. Oliveira, L. D. F. Costa, and M. D. G. V. Nunes, "A complex network approach to text summarization," Information Sciences, vol. 179, no. 5, pp. 584–599, Feb. 2009, doi: 10.1016/j.ins.2008.10.032.
[4] R. M. Aliguliyev, "A new sentence similarity measure and sentence based extractive technique for automatic text summarization," Expert Systems with Applications, vol. 36, no. 4, pp. 7764–7772, May 2009, doi: 10.1016/j.eswa.2008.11.022.
[5] R. M. Alguliev, R. M. Aliguliyev, M. S. Hajirahimova, and C. A. Mehdiyev, "MCMR: Maximum coverage and minimum redundant text summarization model," Expert Systems with Applications, vol. 38, no. 12, pp. 14514–14522, Nov. 2011, doi: 10.1016/j.eswa.2011.05.033.
[6] R. M. Alguliev, R. M. Aliguliyev, and C. A. Mehdiyev, "Sentence selection for generic document summarization using an adaptive differential evolution algorithm," Swarm and Evolutionary Computation, vol. 1, no. 4, pp. 213–222, Dec. 2011, doi: 10.1016/j.swevo.2011.06.006.
[7] Elena Lloret, Manuel Palomar, "Text summarisation in progress: a literature review", Artif Intell Rev (2012) 37:1–41, DOI 10.1007/s10462-011-9216-z.
[8] R. Ferreira et al., "Assessing sentence scoring techniques for extractive text summarization," Expert Systems with Applications, vol. 40, no. 14, pp. 5755–5764, Oct. 2013, doi: 10.1016/j.eswa.2013.04.023.
[9] A. M. Rush, S. Chopra, and J. Weston, "A Neural Attention Model for Abstractive Sentence Summarization." arXiv, Sep. 03, 2015. Accessed: May 30, 2024. [Online]. Available: http://arxiv.org/abs/1509.00685
[10] S. M. Patel, "Extractive Based Automatic Text Summarization," JCP, pp. 550–563, 2017, doi: 10.17706/jcp.12.6.550-563.
[11] Prachi Shah, Nikita P. Desai, "A Survey of Automatic Text Summarization Techniques for Indian and Foreign Languages", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016, 978-1-4673-9939-5/16, IEEE, 2016.
[12] Santosh Kumar Bharti, Korra Sathya Babu, and Sanjay Kumar Jena, "Automatic Keyword Extraction for Text Summarization: A Survey ", arXiv:1704.03242, Feb 2017, https://doi.org/10.48550/arXiv.1704.03242
[13] Riya Kamble, Saurabh Shah, Aalok Nerurkar, Kanhaiya Prasad, Prof. Reena Mahe, "Automatic Text Summarization", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Published by, www.ijert.org, ICIATE - 2017 Conference Proceedings.
[14] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey," Artif Intell Rev, vol. 47, no. 1, pp. 1–66, Jan. 2017, doi: 10.1007/s10462-016-9475-9.
[15] Yike Liu, Tara Safavi, Abhilash Dighe, Danai Koutra, "Graph Summarization Methods and Applications: A Survey", ACM Computing Surveys, Vol. 51, No. 3, Article 62. Publication date: June 2018., https://doi.org/10.1145/3186727.
[16] S. Narayan, S. B. Cohen, and M. Lapata, "Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization." arXiv, Aug. 27, 2018. Accessed: May 30, 2024. [Online]. Available: http://arxiv.org/abs/1808.08745
[17] R. S. Sajjan and M. G. Shinde, "A Detail Survey on Automatic Text Summarization," ijcse, vol. 7, no. 6, pp. 991–998, Jun. 2019, doi: 10.26438/ijcse/v7i6.991998.
[18] D. S. Moirangthem and M. Lee, "Abstractive summarization of long texts by representing multiple compositionalities with temporal hierarchical pointer generator network," Neural Networks, vol. 124, pp. 1–11, Apr. 2020, doi: 10.1016/j.neunet.2019.12.022.
[19] A. Joshi, E. Fidalgo, E. Alegre, and L. Fernández-Robles, "SummCoder: An unsupervised framework for extractive text summarization based on deep auto-encoders," Expert Systems with Applications, vol. 129, pp. 200–215, Sep. 2019, doi: 10.1016/j.eswa.2019.03.045.
[20] C. Yongkiatpanich and D. Wichadakul, "Extractive Text Summarization Using Ontology and Graph-Based Method," in 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore: IEEE, Feb. 2019, pp. 105–110. doi: 10.1109/CCOMS.2019.8821755.
[21] A. K. Mohammad Masum, S. Abujar, M. A. Islam Talukder, A. K. M. S. Azad Rabby, and S. A. Hossain, "Abstractive method of text summarization with sequence to sequence RNNs," in 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India: IEEE, Jul. 2019, pp. 1–5. doi: 10.1109/ICCCNT45670.2019.8944620.
[22] H. Zhang, J. Xu, and J. Wang, "Pretraining-Based Natural Language Generation for Text Summarization." arXiv, Apr. 12, 2019. Accessed: May 30, 2024. [Online]. Available: http://arxiv.org/abs/1902.09243
[23] J. Zhang, K. Li, C. Yao, and Y. Sun, "Event-based summarization method for scientific literature," Pers Ubiquit Comput, vol. 25, no. 6, pp. 959–968, Dec. 2021, doi: 10.1007/s00779-019-01301-5.
[24] S. Murarka and A. Singhal, "Query-based Single Document Summarization using Hybrid Semantic and Graph-based Approach," in 2020 International Conference on Advances in Computing, Communication & Materials (ICACCM), Dehradun, India: IEEE, Aug. 2020, pp. 330–335. doi: 10.1109/ICACCM50413.2020.9212923.

[25] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "EdgeSumm: Graph-based framework for automatic text summarization," Information Processing & Management, vol. 57, no. 6, p. 102264, Nov. 2020, doi: 10.1016/j.ipm.2020.102264.
[26] Rahul, S. Adhikari, and Monika, "NLP based Machine Learning Approaches for Text Summarization," in 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India: IEEE, Mar. 2020, pp. 535–538. doi: 10.1109/ICCMC48092.2020.ICCMC-00099.
[27] A. A. Syed, F. L. Gaol, and T. Matsuo, "A Survey of the State-of-the-Art Models in Neural Abstractive Text Summarization," IEEE Access, vol. 9, pp. 13248–13265, 2021, doi: 10.1109/ACCESS.2021.3052783.
[28] M. M. Haider, Md. A. Hossin, H. R. Mahi, and H. Arif, "Automatic Text Summarization Using Gensim Word2Vec and K-Means Clustering Algorithm," in 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh: IEEE, 2020, pp. 283–286. doi: 10.1109/TENSYMP50017.2020.9230670.
[29] A. Gidiotis and G. Tsoumakas, "A Divide-and-Conquer Approach to the Summarization of Long Documents," IEEE/ACM Trans. Audio Speech Lang. Process., vol. 28, pp. 3029–3040, 2020, doi: 10.1109/TASLP.2020.3037401.
[30] G. Rabby, S. Azad, M. Mahmud, K. Z. Zamli, and M. M. Rahman, "TeKET: a Tree-Based Unsupervised Keyphrase Extraction Technique," Cogn Comput, vol. 12, no. 4, pp. 811–833, Jul. 2020, doi: 10.1007/s12559-019-09706-3.
[31] R. C. Belwal, S. Rai, and A. Gupta, "Text summarization using topic-based vector space model and semantic measure," Information Processing & Management, vol. 58, no. 3, p. 102536, May 2021, doi: 10.1016/j.ipm.2021.102536.
[32] N. Rahman and B. Borah, "Query-Based Extractive Text Summarization Using Sense-Oriented Semantic Relatedness Measure." Nov. 30, 2021. doi: 10.21203/rs.3.rs-1102477/v1.
[33] R. Rani and D. K. Lobiyal, "Document vector embedding based extractive text summarization system for Hindi and English text," Appl Intell, vol. 52, no. 8, pp. 9353–9372, Jun. 2022, doi: 10.1007/s10489-021-02871-9
[34] A. Alomari, N. Idris, A. Q. M. Sabri, and I. Alsmadi, "Deep reinforcement and transfer learning for abstractive text summarization: A review," Computer Speech & Language, vol. 71, p. 101276, Jan. 2022, doi: 10.1016/j.csl.2021.101276.
[35] A. K. Yadav, Ranvijay, R. S. Yadav, and A. K. Maurya, "State-of-the-art approach to extractive text summarization: a comprehensive review," Multimed Tools Appl, vol. 82, no. 19, pp. 29135–29197, Aug. 2023, doi: 10.1007/s11042-023-14613-9.
[36] N.Bhatia, A. Jaiswal,,"Automatic Text Summarizaton and It's Methods-A Review", 978-1-4673-8203-8/16, IEEE 2016
[37] N. Moratanch and S. Chitrakala, "A survey on extractive text summarization," in 2017 International Conference on Computer, Communication and Signal Processing (ICCCSP), Chennai, India: IEEE, Jan. 2017, pp. 1–6. doi: 10.1109/ICCCSP.2017.7944061.
[38] Z. Nasar, S. W. Jaffry, and M. K. Malik, "Textual keyword extraction and summarization: State-of-the-art," Information Processing & Management, vol. 56, no. 6, p. 102088, Nov. 2019, doi: 10.1016/j.ipm.2019.102088.
[39] P. Janjanam and C. P. Reddy, "Text Summarization: An Essential Study," in 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India: IEEE, Feb. 2019, pp. 1–6. doi: 10.1109/ICCIDS.2019.8862030.
[40] N. Baruah, S. Kr. Sarma, and S. Borkotokey, "Text Summarization in Indian Languages: A Critical Review," in 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok, India: IEEE, Feb. 2019, pp. 1–6. doi: 10.1109/ICACCP.2019.8882968.
[41] P. Batra, S. Chaudhary, K. Bhatt, S. Varshney, and S. Verma, "A Review: Abstractive Text Summarization Techniques using NLP," in 2020 International Conference on Advances in Computing, Communication & Materials (ICACCM), Dehradun, India: IEEE, Aug. 2020, pp. 23–28. doi: 10.1109/ICACCM50413.2020.9213079.
[42] A. K. Yadav, A. K. Maurya, Ranvijay, and R. S. Yadav, "Extractive Text Summarization Using Recent Approaches: A Survey," ISI, vol. 26, no. 1, pp. 109–121, Feb. 2021, doi: 10.18280/isi.260112.
[43] D.Manju D, Radhamani V, Dhanush Kannan A, Kavya B, Sangavi S, Swetha Srinivasan, "TEXT SUMMARIZATION", VOLUME 21 : ISSUE 7 (July) - 2022
[44] D. A. K. Chakraverti, A. K. Pandey, A. Dhadse, S. Choudhary, and S. K. Pandey, "A Survey on Methods of Text Summarization," vol. 7, no. 4, 2022.
[45] K. Rathi, Y. V. Singh, and S. Raj, "A Review of state-of-the-art Automatic Text Summarisation," vol. 10, no. 4, 2022
[46] M. Zhang, G. Zhou, W. Yu, N. Huang, and W. Liu, "A Comprehensive Survey of Abstractive Text Summarization Based on Deep Learning," Computational Intelligence and Neuroscience, vol. 2022, pp. 1–21, Aug. 2022, doi: 10.1155/2022/7132226.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)