iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52463

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 11 Issue V May 2023- Available at www.ijraset.com

Deep Learning based Text Abstraction

Mr. Sumit Chougule[1], Mr. Priyansh Dudhabale[2], Mr. Tejas Havaldar[3]
[1, 2, 3]*Department of Electronics and Telecommunication SCTR's Pune Institute of Computer Technology Pune, India*

Abstract: ***Text abstraction based on deep learning has proven to be a promising method for the task of extracting large amounts of text while preserving the most important information. This article provides an overview of text abstraction based on deep learning, highlighting various techniques and applications in this field. This article reviews the existing literature on text abstraction based on deep learning, focusing on various methods such as sentence compression, text summarization, and paraphrase, and compares their advantages and disadvantages. The article also describes various deep learning techniques used in the field, including neural networks, recurrent neural networks, and convolution of neural networks. In addition, this article presents studies on the effectiveness of deep learning-based text in a variety of applications, including journalism, finance, health, and education. The article discusses the challenges faced by the field, such as resolving ambiguity and ensuring consistency and readability in produced texts. Finally, this article discusses future directions and potential areas for further research in deep learning-based text abstraction. This questionnaire is useful for researchers and practitioners interested in text abstraction and applications based on deep learning. The article also explores the ethical implications of deep learning-based reading, particularly with regard to issues such as prejudice and privacy. The benefits of this technology must be weighed against the risks, and it is important to ensure that deep learning-based text is created and used responsibly.***
Keywords: Abstraction, summarization, machine learning.

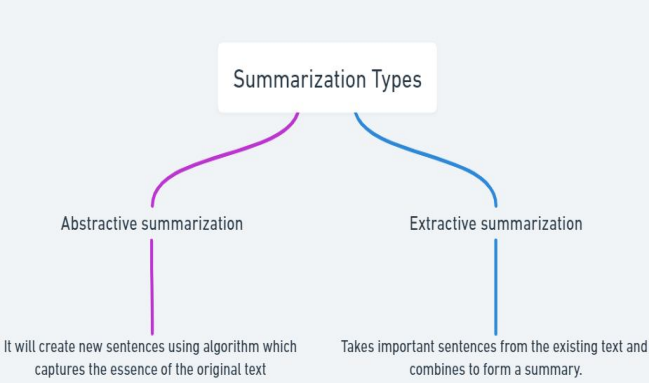
I. INTRODUCTION

Deep learning-based text abstraction is an emerging technique that uses deep learning algorithms to summarize large texts into short, concise summaries while preserving the most important information. This field has grown in importance in recent years due to the explosion of data and information available online, making it difficult for people to process and understand large volumes of text. The purpose of text abstraction based on deep learning is to provide people with a fast and easy way to access and understand complex data. It has many applications in media, finance, health, education and other fields that require the analysis and collection of large amounts of data. It has many techniques such as text abstraction, sentence compression, text writing and annotation based on deep learning. This method uses neural networks, recurrent neural networks and convolutional neural networks to process and analyze the text and generate short text that preserves the content of the original text. The effectiveness of deep learning-based text abstraction has been widely demonstrated in various studies and applications that show good results in terms of accuracy and precision. However, there are still problems in this field such as resolving ambiguous words, ensuring consistency and readability in the produced text. This questionnaire is designed to provide an overview of deep learning based on text abstraction. It will review the available literature on the subject, discuss various applications, and compare and contrast different models. This article also explains the techniques used in text abstraction based on deep learning, presents the findings on the effectiveness of these techniques, and discusses future directions in the field. With the increase of data and information available online, there is a growing need for efficient and comprehensive documentation systems. Text abstraction based on deep learning has emerged as a powerful technique to solve this challenge by producing shorter and shorter versions of longer texts while preserving the core content. Text abstraction based on deep learning includes a number of methods such as summarizing, which involves selecting and combining the most important sentences from the original text, and irony, which includes creating new sentences that capture the essence of the original text. The field has experienced rapid growth in recent years with the availability of big data and the development of deep learning algorithms. Deep learning methods such as convolutional neural networks and recurrent neural networks have been successfully used for text abstraction that can produce high-quality content with high fidelity. Text-based deep learning applications are wide and varied, including text summarization, transcription, and even chatbot responses. For example, copywriting in journalism can help journalists quickly review many stories and extract important information, allowing them to write news accurately and write better stories. Despite significant progress in this area, there are still challenges to be addressed, such as addressing ambiguity, ensuring consistency and readability of output, and reducing the risk of bias or unfair content. However, as technology continues to advance, the benefits of deep learning-based text are overwhelming and are likely to become an important tool in many industries and applications.

©IJRASET: All Rights are Reserved

II. SUMMARIZATION TYPES

There are generally two types of summaries: abstractive summarization and extractive summarization.



Extractive summarization will select the most important sentences or phrases from the original text and combine them to create a summary. This approach preserves the original text of the text and relies on sentence sorting algorithms to identify the most important sentences. Extractive summarization is often used in summarizing news content, where the most important sentences are extracted to create a summary. Abstract summarization will create new sentences that capture the essence of the original text. This approach includes the use of natural language processing and machine learning to understand the meaning of text and create content that is not limited to traditional languages. There is also a third type of summary, called mixed summarization, which combines the elements of extractive summarization and abstractive summarization. Hybrid summarization involves extracting key phrases from the text and then building new sentences based on them to capture the main meaning of the original text. This approach aims to combine the benefits of abstraction and extraction concepts to create interesting and comprehensive content. Extractive summarization will select and combine the most important sentences or phrases from the raw text to create a summary. This method uses statistical techniques and machine learning to identify the most important sentences based on features such as sentence length, word frequency, and relevance. Extractive summarization tends to produce more accurate content as the content is based on existing sentences in the original text. However, since it is based on the original wording of the text, it can sometimes make the content inconsistent or poorly readable. Abstractive summarization will create new sentences that capture the essence of the original text. This approach uses natural language processing techniques and machine learning to understand the meaning of text and create content that is not limited to the original language. Abstract summaries will create a more coherent and legible content as they are not limited to the first word of the text. However, sometimes the generated sentence may not be found in the original text, leading to less incorrect content. Extractive and abstractive summarizations each have advantages and disadvantages and are often used in different contexts and applications. Short summaries are often used for longer documents, such as training documents or technical documents, where the purpose is to make important points and points that retain the main points of the original text.

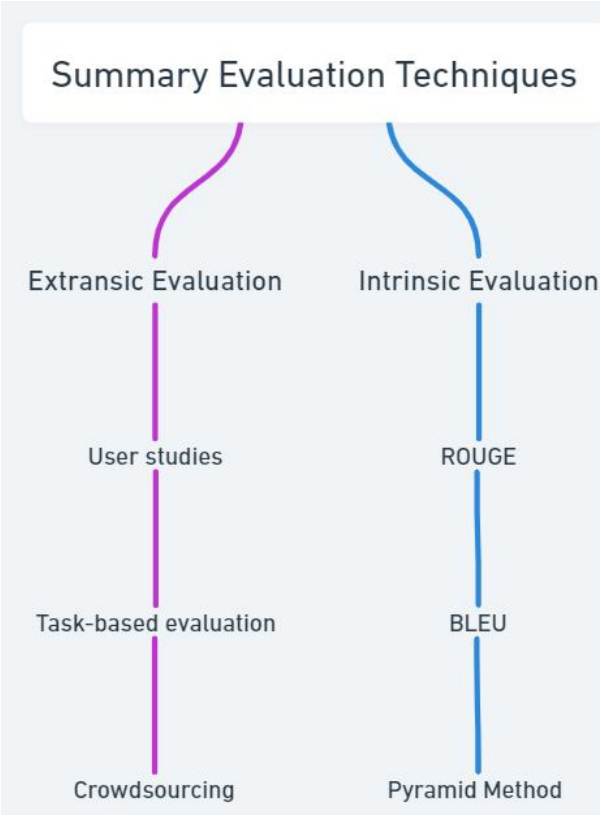
III. SUMMARY EVALUATION TECHNIQUES

Summary evaluation techniques are techniques used to evaluate the quality and effectiveness of content writing. These techniques aim to evaluate whether the abstract design preserves the most important information and key points in the original text and meets the needs and expectations of the target audience. The choice of assessment methods depends on abilities and skills as well as the specific goals and requirements of the written work. Some assessment methods, such as medical or legal texts, may be more appropriate for assessing the quality of abstract content in a particular field, while others may be more appropriate for assessing general terms. Overall, the content evaluation process is important to evaluate the quality and effectiveness of content writing and to guide the improvement of the content writing process, to be accurate and efficient.

The importance of summary evaluation techniques lies in their ability to evaluate the quality of the content produced. Effective review systems are important for improving the accuracy and efficiency of content writing, as they provide feedback on how well the system is performing and where it can be improved. Using the evaluation process, researchers and developers can compare different concepts and methods and identify the advantages and disadvantages of each. This helps guide the development of new content collection methods and improves the overall collection process. The evaluation process is also important for the use of short writing, such as writing a newspaper or writing information for business and research. In these cases, the quality of the content can affect decision making and data processing, so it is important to have a reliable system for evaluating the collection process.

Overall, the importance of the evaluation process lies in their ability to provide objective evaluations that measure the quality of the content produced and guide the development of the article writing process.

Extrinsic and intrinsic evaluation are two different ways of evaluating the effectiveness of content writing.



A. Extrinsic Evaluation

Extrinsic evaluations focus on evaluating the effectiveness of the writing process in the context of a particular project or practice. This includes evaluating whether the content supports the purpose or purpose of the task, such as making a decision or answering a question. External evaluations often include user surveys or functional evaluations; where real users are asked to perform tasks or evaluate content based on its effectiveness and performance.

1) *User studies:* These include real users who perform certain tasks or evaluate abstractions based on their results and performance. User research can provide better insights into how the aggregation meets the needs and expectations of its target audience.
2) *Task-based evaluation:* These involve evaluating the subject's performance in terms of specific tasks or practices, such as providing information or making decisions.
3) *Crowdsourcing:* This involves outsourcing performance appraisals to a number of non-professional evaluators, usually through an online platform. Crowdsourcing is a great way to get feedback from people, but it needs to be carefully organized and managed.

B. Intrinsic Evaluation

Intrinsic evaluation techniques focus on assessing the quality of the abstract, regardless of the context of its use. This includes measuring the consistency, readability and information of the content produced. The intrinsic evaluation process often relies on automated evaluations such as ROUGE and BLEU, which compare produced content with one or more reference content.

1) *Quality evaluation:* The text quality of the summary is checked based on linguistic parameters such as grammar, structure and consistency, vocabulary, and non-duplication.
2) *Informativeness evaluation:* This is the most used type of summary evaluation techniques. There are two ways in which informativeness of summary is evaluated, they are as follows,

Automatic: don't need human annotation
Semi-automatic: needs human annotation some of the informativeness intrinsic evaluation techniques.

C. Rouge

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) makes use of reference summary for evaluation.
It looks for co-occurrences of various levels grams in the generated summary and reference summary. Five different metrics are available to capture ROUGE.
ROUGE-N: checks for overlap of N gram
ROUGE-L: checks for longest common sub-sequences(LCS) ROUGE-W: weighted LCS, favours longest LCS
ROUGE-S: skip-bigram based cooccurrence check
ROUGE-SU: checks of cooccurrence except bi-gram and unigram.

D. Bleu

BLEU (Bilingual Evaluation Understudy)
This is a modified form of precision. The fix is due to overlapping of candidate summaries and link summaries. Here, word overlap in the digest is computed relative to the maximum number of that word among allreference summaries It can be written in the equation as follows,
P = mmax/wt (1)
where mmax is maximum time occurrence of word from all reference summaries and wt is total number of words present in generated summary

E. Basic Element(BE)

Sentences are expressed in the form of using three word namely head, modifier/argument and relation(between head and modifier). Then these are mapped against various equivalence expressions.

F. Depeval

This evaluation method is similar to BE method wherein parsers are used in this method unlike minipar in BE. Dependency triplets (head —modifier— relation) are from the automatically generated text are checked against the ones from reference summaries.

G. Pyramid Method

It is semi-automatic intrinsic informativeness evaluation method which makes use of nation of Summary Content Unit(SCU) which is nothing but the set of sentences with the similar quotient of informativeness. SCUs generated as part of summary and one which are similar to various human level SCUs gets higher weight.

IV. RELATED WORK

According to Shengli Song [1], in 2018, Abstractive Text Summarization (ATS) is concerned with building summary sentences by merging facts from various source sentences and condensing them into short representations while preserving information content and overall meaning. Manually summarising large text documents is very difficult and timeconsuming. In this paper, they propose an LSTM-CNN-based ATS framework (ATSDL) that can construct new sentences by looking at finer-grained fragments than sentences, i.e., semantic phrases. Unlike existing abstraction-based approaches, ATSDL consists of two main phases. The first stage extracts phrases from the source sentences, and the second stage uses deep learning to generate text summaries. Experimental results on the CNN and DailyMail datasets show that the ATSDL framework outperforms state-of-the-art models in terms of both semantics and syntactic structure and achieves competitive results in the manual evaluation of language quality.
Abu Kaisar Mohammad Masum [2] proposed in 2019 that they sequenced models using TensorFlow 1.12.0 sequences. Once training is complete, you can create a custom summary of the machine. Summarization takes input records from the dataset and randomly defines the length of the summary. For the parameters, they used an attention-based encoder. They used the Adam optimizer with epochs = 100, stack size = 64, RNN size = 256, learning rate = 0.005, retention probability = 0.75, and computed the learning rate for each parameter. The vanilla gradient descent optimizer is used to speed up convergence. After data preprocessing, 20% of the data across the data set was retained for testing, and the remaining 80% of the data was provided for training.

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

Testing 4000 data usage with a total of 20k data usage and another training 16000 data. The machine gave maximum positive power. Some response output for some text is wrong. But the positive response output looks incredible.

According to Shristi Rauniyar[3] in 2020. In this paper, they propose various techniques for text summarization, such as Fuzzy C-Means, Deep learning, Machine learning, Transformer, SEN analysis, word embedding, differential methods, graph-based methods, text participation, clustering, Cascade Forest, MOABC. Among these methods, most of them contain extractive summarization and machine learning methods, and we can see that these methods showed better and improved performance than others. From this, we can conclude that the extractive summarization method outperforms the abstractive summarization method and that machine learning techniques are best suited to generate accurate summaries. The goal of their work was to analyze various trends in the field of text-SUMZ.

Ravali Boorugu entitled in 2020 [4] that It's hard to imagine life without smartphones and the internet these days. It has become essential for people of all ages. Along with this increase in internet and smartphone usage, online shopping has also increased steadily. All users check reviews before ordering online. Still, reading one of these long reviews isn't easy for anyone. So, there must be something that can reduce the long review to a short sentence containing his words limited to the same meaning. Text summaries are useful in this aspect. Many of these NLP researchers are interested in his text summary. This document is an overview of various types of text summarization techniques, from basic to advanced. According to this study, the seq2seq model is used in conjunction with LSTM and attention mechanisms to improve accuracy.

According to Dima Suleiman [5], in 2020, the number of online documents has recently increased significantly. Therefore, these documents must be combined to be effective. This paper describes a state-of-the-art approach to extracted text summarization based on deep learning techniques. These approaches fall into three categories based on deep learning techniques: Restricted Boltzmann Machines, Variation Auto-Encoder, and Recurrent Neural Networks. Daily Mail and DUC 2002 are the most commonly used records for extract summaries. Additionally, ROUGE is the primary metric used to assess the quality of the extract aggregation process. The results show that the SummaRunNer approach based on the Gated Recurrent Unit Recurrent Neural Network achieved the highest values in ROUGE1, ROUGE2, and ROUGE-L for Daily Mail records. On the other hand, the approach based on Recurrent Neural Network achieved the best results on the DUC 2002 dataset with terms in ROUGE1 and ROUGE2.

According to Emre DOĞAN[6] in this study, her two categories of sentiment analysis were explored using data collected from many social networks. There is also a topical impression course that summarizes sentences on Twitter. Sentiment analysis was considered a classification problem. To increase the success rate, a study was conducted focusing on words with word embedding methods of semantic context. LSA is used for text summarization. In this study, where both sentiment analysis and text summarization are performed, the main objective is to analyze sentiment and thoughts about the topic and present brief information to the user. The primary model was built using data collected from many social networks. Textual analysis and summaries were created using data from Twitter hashtags. The method used in sentiment analysis was compared to the method of word embeddings and achieved a 93% success rate.

According to Shervin Minaee in 2021 [7], this article provides an overview of over 150 deep learning-based text classification models developed in recent years, discussing their technical contributions, similarities, and strengths. They also provide an overview of over 40 5 common datasets commonly used for text classification. Finally, we provide a quantitative analysis of the performance of various deep learning models against common benchmarks and discuss future research directions.

Nikhil S. Shirwandkar, entitled [8] Approach for Extracted Text Summarization, is designed and implemented for single document summarization. It uses a combination of constrained Boltzmann machines and fuzzy logic to select key phrases from the text but leave a meaningful and lossless summary.

The text document used for the abstract is in English. Various-level sentence and word features are used to provide meaningful sentences.

Two of his summaries are generated for each document using a constrained Boltzmann machine and fuzzy logic. Both summaries are then combined and processed using a series of operations to obtain the final summary of the document. The results show that the designed approach overcomes the problem of text overload by generating effective summaries.

According to K. Yang [9], EcForest is an abstraction summary model with Enhanced Sentence Embedding and Cascade Forest. Sentence representation is very important for many summarization methods. Bags of words can barely capture semantics, and typical embedding models fail to capture more complex semantic features such as ambiguity and phrase meaning. To this end, we propose an Extended Sentence Embedding (ESE) model that solves such shortcomings by mapping multiple valid features into a dense vector. In essence, augmented sentence embeddings are a new model for improving the representation of distributed sentences. They say their sentence embedding model is universal and can be adapted to other NLP tasks.

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

Moreover, Deep Forest is used as a sentence extraction algorithm due to its robustness to hyperparameters and efficient training algorithm compared to deep neural networks. Evaluation of the variant model proposed in this work proves the validity of the improved sentence embeddings. The comparison results of his EcForest with multiple baselines on two different datasets show that the proposed aggregated model outperforms.

According to Shakil Ashraful Anam [10], In this article, they have proposed a sentencebased model that applies an unsupervised classification method, fuzzy CMeans (FCM) clustering, to a sentence ranking method for the purpose of sentence extraction. The sentence 6 rating task relies on five main features, including topic sentences, which are the first novelties in the proposed model. Furthermore, C-means clustering, a soft computing technique typically used for pattern recognition tasks, can be significantly improved by hard clustering the membership of the elements. This was not considered an analogous process in previous studies. To the novelty of the presented model. Using the standard summarization Using the evaluative technique, we measured the accuracy, recognition, and F-measures of the proposed FCM models and compared them from different perspectives with different summarizers. The results show that the FCM model clearly outperforms previous approaches.

Paper Title	Abstract	Technologies
Abstractive text summarization using LSTM-CNN based deep learning. Multimedia Tools and Applications.	In this paper, they propose an LSTM-CNN-based ATS framework (ATSDL) that can construct new sentences by looking at finer-grained fragments than sentences, i.e., semantic phrases. ATSDL consists of two main phases. The first stage extracts phrases from the source sentences, and the second stage uses deep learning to generate text summaries.[1]	LSTM, CNN
Abstractive method of text summarization with sequence-to-sequence RNNs.	In this paper they have used sequenced models using TensorFlow 1.12.0 sequences. Once training is complete, you can create a custom summary of the machine. Summarization takes input records from the dataset and randomly defines the length of the summary. For the parameters, they used an attention-based encoder.[2]	Sequence models
A Survey on Deep Learning based Various Methods Analysis of Text Summarization	In this paper, they propose various techniques for text summarization, such as Fuzzy C-Means, Deep learning, Machine learning, Transformer, SEN analysis, word embedding, differential methods, graph-based methods, text participation, clustering, Cascade Forest, MOABC.[3]	
A Survey on NLP based Text Summarization for Summarizing Product Reviews	This document is an overview of various types of text summarization techniques, from basic to advanced. The seq2seq model is used in conjunction with LSTM and attention mechanisms to improve accuracy.[4]	Sequence to sequence LSTM
Deep learning based extractive text summarization: approaches, datasets and evaluation	This paper describes a state-of-the-art approach to extracted text summarization based on deep learning techniques. These approaches fall into three categories based on deep learning techniques: Restricted Boltzmann Machines, Variation Auto-Encoder, and Recurrent Neural Networks.[5]	RBM,VAE,RNN

measures.		
Deep Learning Based Sentiment Analysis and Text Summarization in Social Networks.	In this study, two categories of sentiment analysis were explored using data collected from many social networks. There is also a topical impression course that summarizes sentences on Twitter. Sentiment analysis was considered a classification problem.[6]	LSTM
Deep learning--based text classification: a comprehensive review.	This article provides an overview of over 150 deep learning-based text classification models developed in recent years, discussing their technical contributions, similarities, and strengths. They also provide an overview of over 40 common datasets commonly used for text classification.[7]	
Extractive text summarization using deep learning.	It uses a combination of constrained Boltzmann machines and fuzzy logic to select key phrases from the text but leave a meaningful and lossless summary.[8]	RBM, Fuzzy logic
EcForest: EXTR document SUMZ through enhanced sentence embedding and cascade forest.	They gave an abstraction summary model with Enhanced Sentence Embedding and Cascade Forest. Sentence representation is very important for many summarization methods. Bags of words can barely capture semantics, and typical embedding models fail to capture more complex semantic features such as ambiguity and phrase meaning. To this end, we propose an Extended Sentence Embedding (ESE) model.[9]	
Automatic text SUMZ using fuzzy C-Means clustering.	In this article, they have proposed a sentence-based model that applies an unsupervised classification method, fuzzy CMeans (FCM) clustering, to a sentence ranking method for the purpose of sentence extraction. The sentence rating task relies on five main features, including topic sentences, which are the first novelties in the proposed model.[10]	fuzzy CMeans clustering

V. CONCLUSION

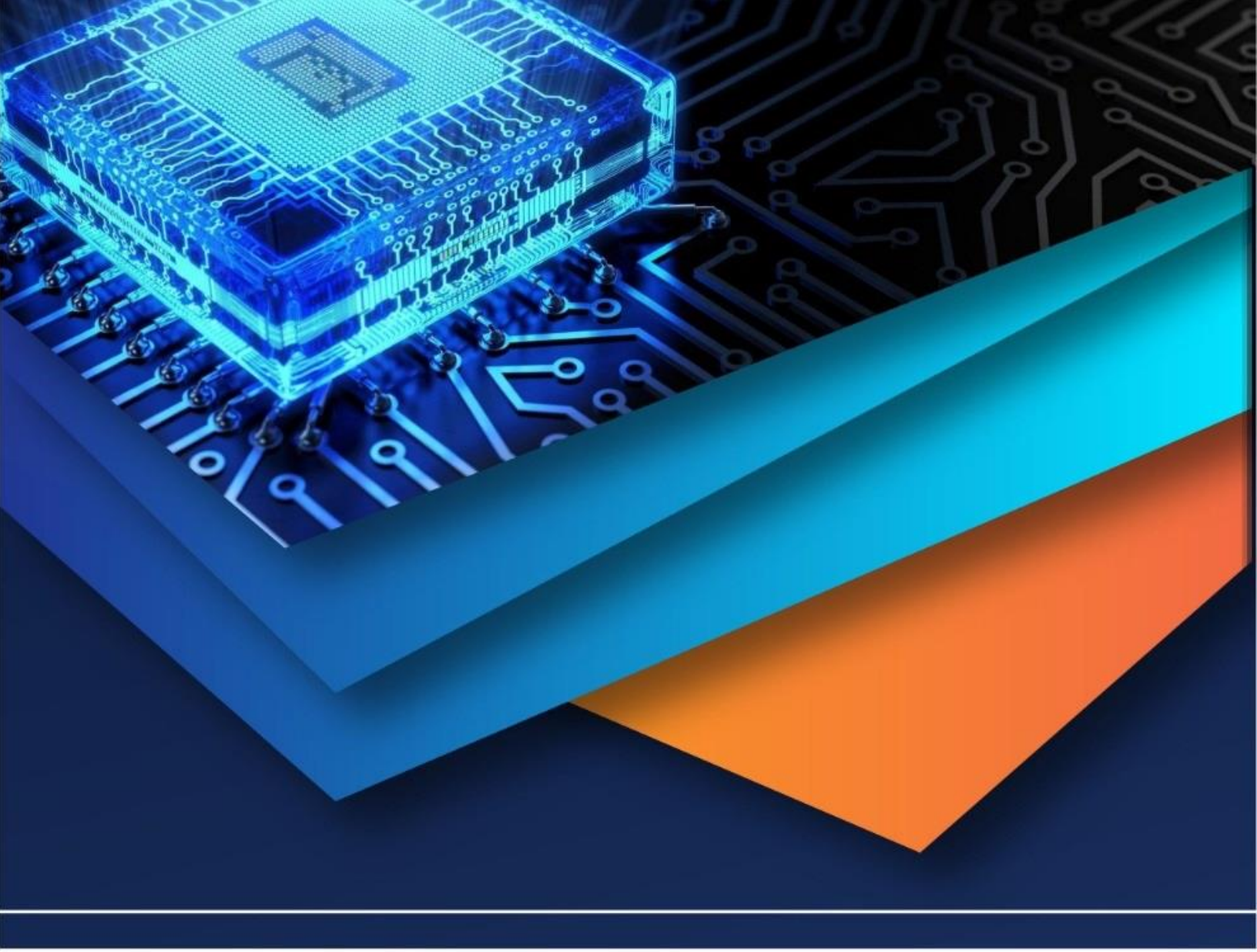
All in all, summary writing using deep learning seems like a good way to write a large number of articles in short and more manageable way. As the value of digital content continues to grow, so does the need for efficient text summarization techniques that can help users quickly extract the most important information from text. In this article, we discuss various techniques and applications in deep learning-based texts, including abstraction and extraction summarization.

We also emphasize the importance of evaluation methods for assessing the quality and effectiveness of written content. There are many challenges and opportunities in text summarization, including the need for more effective writing content and the potential use of aggregators across multiple domains such as content. Text abstraction using deep learning is still a new and rapidly developing field and many exciting developments and advancements are expected in the future. Overall, deep learning-based text analysis has the potential to revolutionize the way we process and extract information from text, helping us manage and understand text about the vast amount of digital content available today. As research and development continues in this area, we can expect to see more efficient and effective writing techniques emerge in the years to come.

REFERENCES

[1] Shengli Song, Shengli, Haitao Huang, and Tongxiao Ruan. "Abstractive text summarization using LSTM-CNN based deep learning." Multimedia Tools and Applications 78.1 (2019): 857-875.
[2] Masum, Abu Kaisar Mohammad, et al. "Abstractive method of text summarization with sequence-to-sequence RNNs." 2019 10th international conference on computing, communication and networking technologies (ICCCNT). IEEE, 2019.
[3] Rahul, S. Rauniyar and Monika, "A Survey on Deep Learning based Various Methods Analysis of Text Summarization," 2020 International Conference on Inventive Computation Technologies (ICICT), 2020, pp. 113-116, doi: 10.1109/ICICT48043.2020.9112474.
[4] R. Boorugu and G. Ramesh, "A Survey on NLP based Text Summarization for Summarizing Product Reviews," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 352-356, doi: 10.1109/ICIRCA48905.2020.9183355.
[5] Suleiman, Dima, and Arafat A. Awajan. "Deep learning based extractive text summarization: approaches, datasets and evaluation measures." 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS). IEEE, 2019.
[6] E. Doğan and B. Kaya, "Deep Learning Based Sentiment Analysis and Text Summarization in Social Networks," 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), 2019, pp. 1-6, doi: 10.1109/IDAP.2019.8875879.
[7] Minaee, Shervin, et al. "Deep learning--based text classification: a comprehensive review." ACM Computing Surveys (CSUR) 54.3 (2021): 1-40.
[8] Shirwandkar, Nikhil S., and Samidha Kulkarni. "Extractive text summarization using deep learning." 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE, 2018.
[9] K. Yang, H. He, K. Al-Sabahi, and Z. Zhang, “EcForest: EXTR document SUMZ through enhanced sentence embedding and cascade forest,” Concurr. Comput. , no. May 2018, pp. 1–12, 2019.
[10] S. A. Anam, A. M. Muntasir Rahman, N. N. Saleheen, and H. Arif, “Automatic text SUMZ using fuzzy C-Means clustering,” 2018 Jt. 7th Int. Conf. Informatics, Electron. Vis. 2nd Int. Conf. Imaging, Vis. Pattern Recognition, ICIEV-IVPR 2018, pp. 180– 184, 2019

©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)