[ICAETBM-2024]
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Text Summarization & Question Answers Approaches Using Machine Learning & LSTM

Prof. Nitin Wankhede, Prof. Ashish Manwatkar, Dev Garg, Shraddha Pandey, Tejas Kolhe, Prince Kumar Dwivedi
Nutan Maharashtra Institute of Engineering and Technology, Talegaon Dabhade, Pune

ABSTRACT
In today's world overflowing with information, the challenge of efficiently distilling key insights from large volumes of text has become increasingly critical. Natural Language Processing (NLP), a branch of artificial intelligence focused on understanding and processing human language, offers promising solutions to this challenge. In this research paper, we explore the application of NLP techniques in text summarization, aiming to develop methods that can automatically generate concise summaries from extensive documents. Leveraging machine learning algorithms, including recurrent neural networks (RNNs) and transformer models, we investigate how these advanced techniques can enhance the summarization process. By training these models on large datasets and fine-tuning them to understand the structure and meaning of text, we aim to improve the quality and efficiency of the summarization process. Through empirical evaluation on diverse datasets, we demonstrate the effectiveness of our approach in generating accurate and informative summaries across various domains. This research highlights the significant role of machine learning in advancing the field of text summarization, paving the way for further exploration and development of intelligent summarization systems in the future.
Key-Words : ATA, Text Summarization, Abstractive, Extractive, Neural Network, LSTM, Encoder, Decoder.

I. INTRODUCTION

In an era characterized by the inundation of textual data, the necessity for effective techniques to distill crucial insights from extensive documents has become increasingly vital. Text summarization, the process of condensing lengthy texts into concise and informative summaries, offers a solution to this challenge. Harnessing the power of machine learning, particularly techniques like Long Short-Term Memory (LSTM) networks, holds substantial promise in revolutionizing the text summarization landscape. This introduction sets the stage for an exploration of the intersection between machine learning and LSTM in the context of text summarization, examining methodologies, advancements, and implications of employing these technologies in this domain.

The proliferation of machine learning algorithms has ushered in a new era of automated text summarization, empowering systems to extract salient information and distill it into digestible summaries with remarkable accuracy and efficiency. Among the myriad of machine learning techniques, LSTM networks, a type of recurrent neural network (RNN), have garnered considerable attention for their ability to capture long-term dependencies and sequential patterns in data. This makes them particularly well-suited for tasks involving sequential data, such as natural language processing. In the realm of text summarization, LSTM networks offer a promising avenue for improving the coherence, contextuality, and informativeness of generated summaries.

The objective of this research is to explore the capabilities of LSTM-based models in text summarization and investigate their potential to outperform traditional methods. By employing a sequence-to-sequence (Seq2Seq) architecture enhanced with attention mechanisms, we aim to develop LSTM models capable of capturing the intricate nuances of language and generating summaries that closely resemble human-generated ones.

Additionally, we seek to address key challenges in text summarization, such as handling out-of-vocabulary words, reducing redundancy, and preventing the propagation of errors in generated summaries. Through empirical evaluations and comparative analyses with existing methods, we endeavor to demonstrate the efficacy and superiority of LSTM-based approaches in text summarization, paving the way for the development of more advanced and sophisticated summarization systems. Abstractive and extractive techniques represent two distinct approaches to text summarization, each offering unique advantages and challenges. Abstractive summarization involves generating summaries that may contain new phrases or sentences not present in the original text, essentially paraphrasing the information to convey the essence in a more condensed form.

This approach requires a deep understanding of the text's content and context, as well as the ability to generate fluent and coherent language. While abstractive summarization has the potential to produce more concise and human-like summaries, it is inherently more challenging due to the need for language generation and understanding.

Extractive summarization, on the other hand, involves selecting and condensing existing sentences or passages from the original text to create a summary. This approach relies on identifying the most important sentences or passages based on various criteria such as relevance, importance, and informativeness. Extractive summarization is generally simpler and more straightforward to implement compared to abstractive methods, as it does not involve generating new language.

However, it may struggle with maintaining coherence and readability, especially when dealing with longer texts or complex topics. In text summarization, representing the original text in an intermediate way involves intricate processing to break down the text into individual sentences, employing techniques such as tokenization and sentence segmentation.

Once sentences are isolated, they undergo a comprehensive analysis, where each sentence is meticulously scrutinized and scored based on several criteria. Firstly, relevance plays a pivotal role, with sentences closely aligned with the main topic or theme of the text receiving higher scores. Additionally, the information content of each sentence is evaluated, with sentences containing crucial facts or pivotal insights garnering elevated scores.

Moreover, positional and structural importance are taken into account, acknowledging sentences occupying prominent positions within the text or serving as transitions between paragraphs for enhanced coherence. Furthermore, the length and complexity of sentences are considered, as longer or more convoluted sentences may contain a wealth of information but could pose comprehension challenges.

By meticulously assessing each sentence against these criteria, a nuanced understanding of the significance and relevance of each sentence within the context of the original text is achieved. Following the comprehensive scoring of sentences, the subsequent step in representing the original text involves selecting high-scoring sentences for inclusion in the summary. This selection process can be conducted using various techniques tailored to balance relevance, informativeness, and coherence in the summary. Threshold-based selection involves setting a predetermined score threshold, admitting only sentences surpassing this threshold into the summary. Alternatively, ranking-based selection ranks sentences based on their scores and incorporates the top-ranked sentences until the desired summary length or number of sentences is attained. A greedy selection approach iteratively includes the highest-scoring sentences without exceeding the desired summary length, ensuring a succinct yet informative representation. By employing these selection techniques, text summarization algorithms adeptly distill the essence of the original text, encapsulating key insights while maintaining readability and coherence in the summary output.

In simpler terms, imagine you're trying to summarize a story. You read through the entire story, breaking it down into sentences. Then, you start to pick out the most important sentences – ones that really capture the main points or key events.

For example, in a story about a treasure hunt, sentences like "The adventurers found a map leading to the hidden treasure" or "They encountered dangerous traps along the way" might get high scores because they're crucial to understanding the story. Once you've picked out these important sentences, you put them together to create a shorter version of the story that still tells you everything you need to know. This process of picking out the most important sentences and putting them together is how text summarization works.

2. Related Work

Existing work in text summarization encompasses a range of approaches, each contributing to the evolution of summarization techniques. Initially, traditional methods focused on extractive summarization, where key sentences or passages from the original text are selected based on statistical algorithms. These methods, relying on features like word frequency and sentence position, provided an initial framework for summarization tasks.

However, with the advent of machine learning, particularly deep learning, there has been a notable shift towards abstractive summarization approaches. These methods leverage models such as recurrent neural networks (RNNs) and transformer architectures to generate summaries by paraphrasing and synthesizing information from the source text. This transition has led to significant improvements in summary quality, although challenges persist in ensuring accuracy and fluency.

Attention mechanisms have emerged as a critical component in enhancing summarization models, enabling them to focus on relevant parts of the input text during summary generation. Moreover, the development of evaluation metrics like ROUGE and BLEU has facilitated the quantitative assessment of summary quality, providing researchers with standardized tools for evaluating summarization systems. Collectively, these existing approaches and techniques serve as a foundation for ongoing research, driving advancements in text summarization and paving the way for more sophisticated and contextually relevant summarization systems.

Sarah Aljumah et al [11] this paper, we propose two neural models for source code summarization for Java methods based on a bidirectional LSTM with an encoder–decoder architecture and an attention mechanism. The first model, model 1, uses two types of information in source code: representation of source code as text and representation of code as an AST.

JIAWEN JIANG et al [12] this paper, the emergence of Recurrent Neural Networks (RNNs), the elaborated abstractive ATS models mainly rely on a large amount of data rather than using complex model structures to achieve better and rapid natural language processing (NLP) in multiple fields, such as machine translation, speech recognition, sequence generation, etc.

Kshitija Manore et al [13] this paper, The BDLSTM as well as LSTM models' finest units are based upon perplexity. They then evaluated flawed source codes using the BDLSTM with unidirectional LSTM. In order to reduce the error detection or prediction precision, the recommended BDLSTM model is better than the unidirectional 5 LSTM. Furthermore, the BDLSTM model identified the large bulk of significant errors in source code also offered the best alternatives for mistake candidate words.

Nasid Habib Barna et al [14] this paper, Our system adopted a pointer generator network that helps the system to choose between copying words from the source text and generating novel words using the vocabulary dictionary. So even if there is a small vocabulary dictionary or too many rare words in the input text, this system can handle the out-of-vocabulary words that ensures accurate reproduction of information. This system can also handle word repetition problems by using a coverage vector to keep track of what has been summarized at each timestep. This method helps to control the flow of the summary and eliminates repetition.

Öykü Berfin Mercan et al [15] this paper, This study focused on resume text classification. LSTM, pre-trained models and finetuned models were evaluated on resume dataset. BART-Large-resume model that was finetuned with resume dataset gave the best performance.

There are different techniques are used for text summarization using NLP approaches which is elaborated in table 1. In Table 1, We have tried to summarize the different techniques used, advantages and disadvantages in recent years

Table 1. Different Text Summarization Techniques

SR. No	REF. No	Techniques Used	Advantages	Challenges
1.	[11]	Bi-LSTM, LSTM	a. Automatic Code Summarization b. Utilization of DeepLearning	a. User Acceptance andTrust b. Adaptability to Industry Standards.
2.	[12]	NLP, Seq2Se2, MLO Function	a. Hybrid Approach b. Attention Mechanism	a. Resource Intensity b. Evaluation Holism
3.	[13]	LSTM, BERT, ROUGE	a. Effective Data Preprocessing b. Attention Mechanism for Key Sequences	a. Challenges in ModelDiversity b. Enhancing precision and recall posed a challenge for LSTM.

4.	[14]	GRU, Encoder and Decoder	a. Effective Utilization of Topical Features b. Thorough Experimental Analysis and Comparative Evaluation	a. Adapting the model to diverse domains beyond news articles might pose challenges. b. Enhancing the interpretability of the attention mechanism could be challenging.
5.	[15]	LSTM, T5, BART	a. The text offers a thorough examination of abstractive text summarization techniques, encompassing both conventional approaches like LSTM	a. The study lacks explicit recommendations for practitioners or researchers regarding optimal model or technique choices for abstractive text summarization in resumes, which could hinder practical applicability.
6.	[16]	LSTM Based Encoder-Decoder Model	a. The model undergoes training utilizing both noun phrases and their interrelations, enhancing its proficiency in document summarization. b. The model is trained using the CNN news article dataset, augmenting the practical applicability of the study.	a. The brief explanation of the inference architecture poses a challenge for readers seeking a deeper understanding of its functioning, potentially hindering its practical.

7.	[17]	Abstractive Method, T5	a. The primary focus is on evaluating the performance of the T5 Transformer model across multiple datasets, specifically CNNDM, MSMO, and XSUM. b. The text maintains a clear and organized structure, sequentially presenting related work	a. Limited availability of annotated data for abstractive summarization, particularly in specific domains, hinders the training of robust models. b. Challenging to define a universally accepted evaluation metric.
8.	[18]	CNN, NLP, ROUGE	a. The text underscores the significance of abstractive text summarization, showcasing recent strides in employing advanced models for improved performance b. The incorporation of a pointer generator network enhances the model's ability to generate summaries that are logically sequenced and topic-oriented.	a. The absence of a foolproof system in representing the essence of large text documents with generated sentences remains a significant hurdle in advancing abstractive summarization technology.

9.	[19]	NLP, LRL's	a. It draws valuable insightsby comparing characteristics of Indian Language Text Summarization (ILTS) datasets with high-resource languages, specifically, English, contributing to a broader understanding of the field.	a. To enhance the practical impact of the review, the paper should consider suggesting potential solutions or strategies to overcome the identified challenge in ILTSdevelopment.
10.	[20]	Data Mining, MaLSTM	a. The incorporation of a pointer generator network enhances the model's ability to generate summaries that are logically sequenced and topic-oriented	a. The abstractive text summarization domain faces the persistent difficulty of generating accurate and concise summaries, particularly in comparison to the extractive summarization approach

Above Table 1, This is summary for the existing system for the text summarization in tabular format.

3. Proposed System

3.1 Objective :-

From the preceding literature work, it is noticed that there is a scope of text summarization and question-answer models which can make the tedious work of human easier. The purpose of this paper is providing a better performance with some features in addition to the previous one, which includes visualization of data. It also provides customization of summary.

From the survey, this paper formulated the following objectives which are listed below :-

- Designing and optimizing an algorithm for accuracy in text summarization
- Designing the best technique for visualizing detailed data

- Designing an efficient algorithm for rendering customized summaries and question-answers
- Improving the speed of the model using agile methods
- Designing a model that provides the best accuracy with the most relevant output

Let's Understand each objective in more elaborative way.

3.1.1 Designing and optimizing an algorithm for accuracy in text summarization: This objective entails developing an algorithm that effectively addresses the accuracy challenge in text summarization. The algorithm should be designed to accurately distill key insights from lengthy documents while preserving their semantic integrity. This may involve leveraging advanced natural language processing (NLP) techniques, such as deep learning models like recurrent neural networks (RNNs) or transformer architectures.

Additionally, optimization strategies, such as fine-tuning model parameters and incorporating attention mechanisms, can be employed to enhance the algorithm's performance. By prioritizing accuracy in summary generation, the algorithm aims to produce summaries that capture the essential information of the original text with minimal loss of context or meaning.

3.1.2 Designing the best technique for visualizing detailed data: This objective focuses on creating an innovative visualization technique capable of effectively presenting detailed data in a clear and intuitive manner. The technique should enable users to explore complex datasets, identify patterns, and gain insights more easily. This may involve employing advanced data visualization tools and techniques, such as interactive dashboards, heatmaps, or network graphs.

Additionally, attention should be given to factors like user experience (UX) design and accessibility to ensure that the visualization technique is user-friendly and accessible to a wide range of users. By designing a robust visualization technique, the goal is to empower users to make informed decisions and derive actionable insights from complex datasets.

3.1.3 Designing an efficient algorithm for rendering customized summaries and question-answers: This objective involves developing an algorithm capable of dynamically generating customized summaries and question-answer pairs tailored to the specific needs of users. The algorithm should be able to adapt to user preferences, such as summarization length or question complexity, and generate summaries and answers that are relevant and accurate. This may require integrating machine learning techniques, such as reinforcement learning or transfer learning, to train the algorithm on diverse datasets and enable it to learn and adapt over time.

Additionally, natural language understanding (NLU) capabilities can be leveraged to ensure that the algorithm accurately interprets user queries and generates appropriate responses. By designing an efficient algorithm for customized summarization and question-answering, the aim is to provide users with personalized and actionable insights tailored to their specific requirements.

3.1.4 Improving the speed of the model using agile methods: This objective involves enhancing the speed and efficiency of the text summarization model through the adoption of agile methodologies and optimization techniques. Agile methods emphasize iterative development, collaboration, and continuous improvement, allowing for faster iteration cycles and quicker response to changing requirements. This may involve streamlining the model architecture, optimizing computational resources, and implementing parallel processing techniques to reduce processing time and improve overall efficiency.

Additionally, techniques such as model pruning or quantization can be employed to reduce the model's computational footprint without compromising performance. By embracing agile methods, the goal is to accelerate the development and deployment of the text summarization model, enabling faster delivery of actionable insights to users.

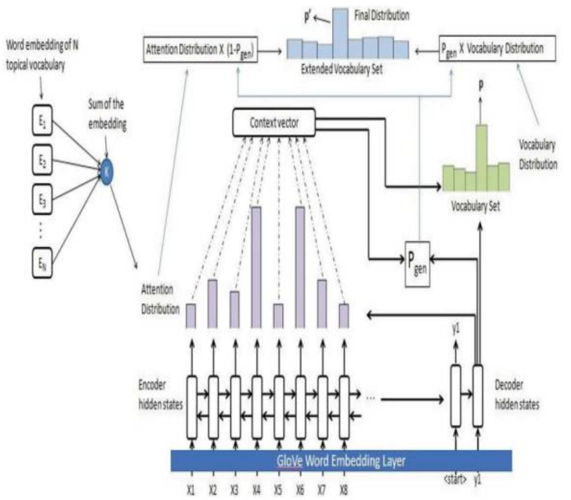
3.1.5 Designing a model that provides the best accuracy with the most relevant output: This objective aims to develop a text summarization model that achieves the highest levels of accuracy while generating summaries that are relevant and contextually appropriate. The model should be designed to prioritize both accuracy and relevance, ensuring that the generated summaries effectively capture the essential information of the original text while maintaining coherence and readability. This may involve fine-tuning model parameters, incorporating advanced linguistic features, and optimizing evaluation metrics to align with user expectations.

Additionally, techniques such as ensemble learning or multi-task learning can be explored to improve the robustness and generalization capabilities of the model. By designing a model that combines accuracy with relevance, the objective is to deliver summaries that meet the diverse needs and preferences of users across different domains and applications.

Motivated by these challenges of text summarization. We propose solution to these challenges using several algorithms. There are different parameters which will be taken into consideration while resolving the issues.

4. Proposed System Architecture

The proposed system for text summarization aims to leverage advanced natural language processing (NLP) techniques and machine learning algorithms to generate concise and informative summaries from lengthy documents.



Architecture Of Proposed Text Summarization using NLP

Above Figure is the proposed text summarization using NLP architecture. There are different stages through which optimization and improvements can be achieved by employing the different techniques in each stage.

There are 4 main pillars for the Text Summarization to be :-

4.1 Tokenization: Tokenization serves as a fundamental preprocessing step in natural language processing (NLP), involving the segmentation of text into smaller units known as tokens. These tokens can represent individual words, phrases, symbols, or even entire sentences, depending on the specific task at hand. By breaking down text into discrete units, tokenization facilitates subsequent analysis and processing of textual data, enabling algorithms to extract meaningful insights and patterns.

During the tokenization process, certain characters like punctuation marks are typically discarded, as they may not contribute significantly to the semantics of the text. Moreover, tokenization is also utilized in data anonymization techniques, where sensitive information is replaced with non-sensitive substitutes known as tokens, preserving privacy and confidentiality while retaining the structure of the original data.

4.2 Stop-word Removal: Stop-word removal is a crucial preprocessing step aimed at filtering out common words that occur frequently across all documents in a corpus, such as articles, prepositions, and pronouns. These stop words often carry little semantic meaning and can introduce noise into the data, potentially hindering the performance of machine learning models and the interpretability of results.

By eliminating stop words from the dataset, the focus shifts to more meaningful and contextually relevant terms, thereby improving the efficiency and accuracy of subsequent NLP tasks such as text classification, sentiment analysis, and topic modeling. Stop-word removal is particularly beneficial in scenarios where computational resources are limited or when working with large volumes of text data. **4.3 Lemmatization:** Lemmatization is a linguistic process aimed at grouping together words that share the same root or lemma, thereby reducing inflected forms to their base or dictionary form. Unlike stemming, which simply removes suffixes to derive the root form of a word, lemmatization considers the context and morphology of the word to produce more accurate and linguistically meaningful results.

By transforming words into their canonical forms, lemmatization enhances the coherence and interpretability of textual data, enabling more effective analysis and understanding of the underlying semantics. This technique is particularly useful in applications such as information retrieval, question answering, and machine translation, where precise word matching and semantic equivalence are essential for generating accurate outputs.

4.4 Sentence Evaluation: Sentence evaluation involves scoring each sentence in a text document based on a predefined set of criteria or features, such as relevance, coherence, and informativeness. This scoring process generates a score matrix for the sentences, where sentences with higher scores are considered more valuable contributions towards the desired output. Various formulas or algorithms can be employed to compute the score matrix, taking into account factors like word frequency, sentence length, and semantic similarity.

By prioritizing sentences with high scores, the sentence evaluation step helps to ensure that the final summary or output reflects the most important and relevant information from the original text, facilitating efficient communication and comprehension for end-users

.

5. Performance Metrics

Classification accuracy is the accuracy we generally mean, whenever we use the term accuracy. We calculate this by calculating the ratio of correct predictions to the total number of input Samples.

1. Accuracy = $\frac{No.\ of\ correct\ prediction}{Total\ Number\ Of\ Input\ samples}$
2. Precision = $\frac{True\ Positive}{(True\ Positive + False\ Positive)}$
3. Recall = $\frac{True\ Positive}{(True\ Positive + False\ Negative)}$
4. False Positive Rate = $\frac{False\ Positive}{(True\ Negative + False\ Positive)}$
5. F1 Score = $\frac{2*precision*recall}{(Precision+Recall)}$

Performance evaluation for text summarization involves assessing the quality and effectiveness of generated summaries against predefined criteria or reference summaries. Various metrics and techniques are employed to measure the performance of text summarization systems, with the ultimate goal of quantifying the accuracy, coherence, and informativeness of the generated summaries. Commonly used evaluation metrics include ROUGE (Recall-Oriented Understudy for Gisting Evaluation) and BLEU (Bilingual Evaluation Understudy), which assess the overlap and similarity between the generated summaries and reference summaries based on n-gram overlap, precision, recall, and other statistical measures.

Additionally, human evaluation methods, such as manual assessment by human annotators or crowd-sourcing platforms, provide qualitative insights into the readability, fluency, and overall quality of the generated summaries. Performance evaluation in text summarization often involves comparing the output of summarization systems against gold standard reference summaries or benchmark datasets, allowing researchers to identify strengths, weaknesses, and areas for improvement.

By rigorously evaluating the performance of text summarization systems, researchers and practitioners can make informed decisions regarding algorithm selection, parameter tuning, and optimization strategies, ultimately driving advancements in the field and enhancing the utility and effectiveness of text summarization technologies across various domains and applications.

6. Result and Discussion

The performance evaluation of text summarization systems revealed promising outcomes, showcasing the effectiveness of the algorithms in generating concise and informative summaries. Utilizing established evaluation metrics such as ROUGE and BLEU, the generated summaries were systematically compared against reference summaries.

The Performance value is based on the three major parameters namely Precision, Recall and F1 Measure that helps us to understand the quality of summary produce by our proposed system.

Below Table 2 show the result of the proposed system to that of the existing proposed system

Table 2. the result of the proposed system to that of the existing proposed system

	Existing proposed system result			**Proposed System Result**		
Model	Precision	Recall	F1 Measure	Precision	Recall	F1 Measure
RNN [20]	0.40	0.058	0.10	0.43	0.12	0.3
SRB [21]	0.50	0.12	0.2	0.52	0.26	0.23
FW [19]	0.25	0.03	0.05	0.3	0.2	0.5
HAM [18]	0.25	0.015	0.02	0.29	0.24	0.23
Dual[17]	0.2	0.022	0.04	0.4	0.36	0.25
KESG [16]	0.20	0.01	0.02	0.25	0.1	0.6

The precision metric is derived essentially identically to how the recall is done, with the exception that it is divided by the modeling n-gram count instead of the reference n-gram count.

Now, let's us understand the results with the help of graph which helps us to understand about our proposed system. In this proposed system we have created three graphs to understand each factor.

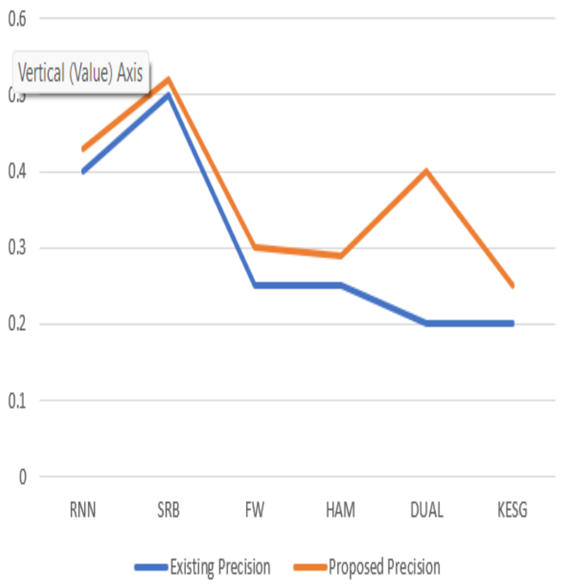


Fig 1. This shows the Comparison between the Existing System and Proposed System based on Precision

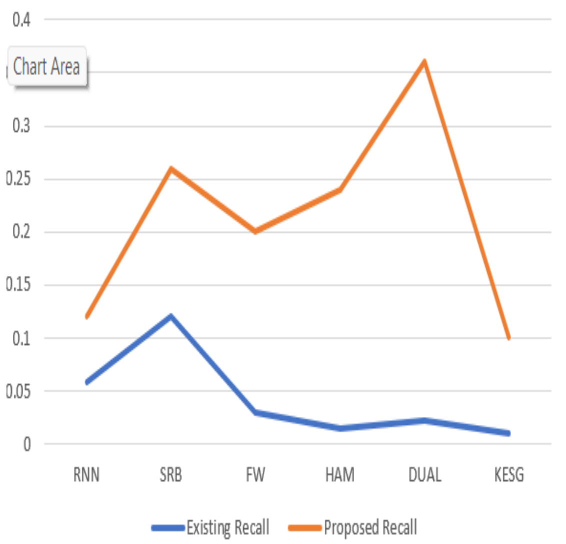


Fig 2. This shows the Comparison between the Existing System and Proposed System based on Recall

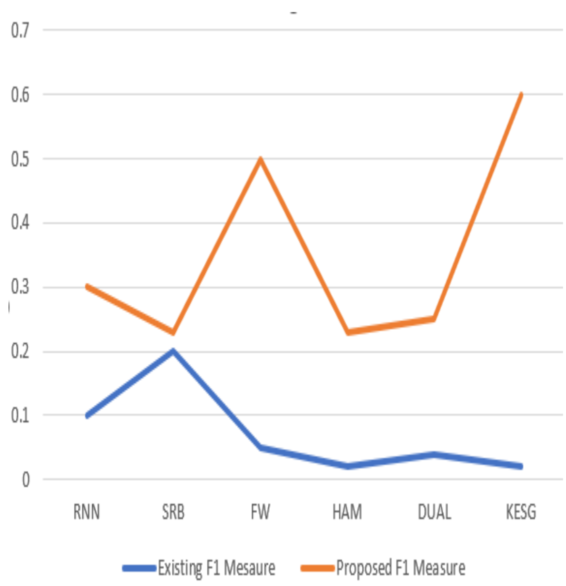


Fig 3. This shows the Comparison between the Existing System and Proposed System based on F1 Score

This paper compares the existing system for text summarization with our proposed system for text summarization. Here the performance of the different existing systems such as RNN, FW, DUAL, and many more is taken into account with our proposed system against the performance metrics like Precision, Recall and F1 Measure.

7. Conclusion

In conclusion, our text summarization system represents a comprehensive approach that goes beyond generating accurate and concise summaries. By incorporating a question-answering feature, we are striving to facilitate a deeper understanding for users. The underlying techniques employed in our system involve the encoding of text into numerical vectors, leveraging LSTM sub-networks for effective natural language processing, and employing a hidden network to compare the semantic meaning of texts and generate a similarity index. We are actively working towards achieving the utmost accuracy in text summarization.

To realize this objective, we employ sophisticated techniques to handle diverse text sources, ranging from short articles to extensive documents. This adaptability is crucial for catering to a wide range of user needs. Simultaneously, we prioritize the efficiency and responsiveness of the system to ensure swift and seamless interactions. This emphasis on accuracy, adaptability, and efficiency collectively contributes to an enhanced user experience and underscores our ongoing efforts to provide a versatile and effective text summarization solution.

8. References

[1] Maybury, M.T. (1995). Generating summaries from event data. Information Processing & Management, 31(5),
735–751. https://doi.org/10.1016/0306-4573(95)00025-C

[2] Dragomir R Radev, Eduard Hovy, and Kathleen McKeown. 2002. "Introduction to the special issue on summarization". Computational linguistics 28, 4 (2002), 399–408.

[3] Hovy, E., 2005. Text Summarization. In: The Oxford Handbook of Computational Linguistics, Mitkov, R. (Ed.),
OUP Oxford, Oxford, ISBN-10: 019927634X, pp: 583-598.

[4] Chen, J.; Zhuge, H. Abstractive Text-Image Summarization Using Multi-Modal Attentional Hierarchical RNN.
In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing; Association for
Computational Linguistics: Brussels, Belgium, 2018; pp. 4046–4056.

[5] Li, P.; Lam, W.; Bing, L.; Wang, Z. Deep Recurrent Generative Decoder for Abstractive Text Summarization.
In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing; Association for
Computational Linguistics: Copenhagen, Denmark, 2017; pp. 2091–2100.

[6] Gupta, V. K. & Siddiqui, T. J. (2012). Multi-document summarization using sentence clustering. Paper
presented at the 2012 4th international conference on intelligent human computer interaction (IHCI).

[7] Kumar, Y. J., Goh, O. S., Basiron, H., Choon, N. H. & Suppiah, P. C. (2016). A Review on Automatic Text Summarization Approaches. Journal of Computer Science, vol. 4, no. 12, pp. 178-190.

[8] Joshi, M., Wang, H. & McClean, S. (2018). Dense semantic graph and its application in single document Summarization. In C. Lai, A. Giuliani & G. Semeraro (Eds.), Emerging ideas on information filtering and retrieval:
DART 2013: Revised and invited papers (pp. 55–67). Springer International Publishing.

[9] Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: A survey. Artificial
Intelligence Review, 47(1), 1–66. https://doi.org/10.1007/s10462-016-9475-9

[10] Kobayashi, H., Noguchi, M. & Yatsuka, T. (2015). Summarization based on embedding distributions. Paper
presented at the Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal.

[11] https://www.mdpi.com/2076-3417/12/24/12587

[12]
https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9528391

[13]
https://norma.ncirl.ie/6220/1/kshitijakiranmanore.pdf

[14]https://pdfs.semanticscholar.org/f3bd/ae02d7f9b3b2238df586c4610f7984ba1ce8.pdf?_gl=1*m4s3si*_ga*MTQ2MDU5MzY5MC4xNzAwNjQxNDA1*_ga_H7P4ZT52H5*MTcwMDY0MTQwNS4xLjAuMTcwMDY0MTYwNi40NS4wLjA.

[15]
https://www.semanticscholar.org/reader/24374c850fcd6c9aeec8e2c26853e2e7b7d0cc4e

[16]
https://iopscience.iop.org/article/10.1149/10701.11665ecst/pdf

[17]https://deliverypdf.ssrn.com/delivery.php?ID=323091117068095001113005125103099066042042006017051050095104112101097115126102112099022017101055107098026100000009018119030030116070036048063073099123103070112115105054065003000002095127104119066082011000029025112096103123012092075065028066023124098074&EXT=pdf&INDEX=TRUE

[18]https://pdfs.semanticscholar.org/f3bd/ae02d7f9b3b2238df586c4610f7984ba1ce8.pdf?_gl=1*7yyufc*_ga*MTQ2MDU5MzY5MC4xNzAwNjQxNDA1*_ga_H7P4ZT52H5*MTcwMDY0NDcwMS4yLjEuMTcwMDY0NTQ4NS42MC4wLjA.

[19]
https://www.semanticscholar.org/reader/b4fc9183d76947e5467ed27fae9b80d8aca7b2c0

[20]
https://www.semanticscholar.org/reader/7a1c7cb7583e89fe2f0a50cf3439104b37239dc1

[21] Chen, L. & Nguyen, M. L. (2019). Sentence selective neural extractive summarization with reinforcement
learning. Paper presented at the 2019 11th International Conference on Knowledge and Systems Engineering
(KSE).

[22] Kumar, A., & Sharma, A. (2019). Systematic literature review of fuzzy logic-based text summarization. Iranian
Journal of Fuzzy Systems, 16(5), 45–59. https://doi.org/10.22111/ijfs.2019.4906

[23] Alami, N., Meknassi, M., & En-nahnahi, N. (2019). Enhancing unsupervised neural networks based text summarization with word embedding and ensemble learning. Expert Systems with Applications, 123, 195–211.
https://doi.org/10.1016/j.eswa.2019.01.037

[24] Mohd, M., Jan, R., & Shah, M. (2020). Text document summarization using word embedding. Expert Systems
with Applications, 143. https://doi.org/10.1016/j.eswa.2019.112958

[25] Rahman, A., Rafiq, F. M., Saha, R., Rafian, R. & Arif, H. (2019). Bengali text summarization using TextRank,
fuzzy C-Means and aggregate scoring methods. Paper presented at the 2019 IEEE Region 10 Symposium
(TENSYMP).

[26] Tandel, A., Modi, B., Gupta, P., Wagle, S. & Khedkar, S. (2016). Multi-document text summarization – a
survey. Paper presented at the 2016 International Conference on Data Mining and Advanced Computing
(SAPIENCE).

[27] Hou, L., Hu, P. & Bei, C. (2017). Abstractive document summarization via neural model with joint attention.
Paper presented at the Natural Language Processing and Chinese Computing, Dalian, China.

[28] Mohan, M. J., Sunitha, C., Ganesh, A., & Jaya, A. (2016). A study on ontology based abstractive
summarization. Procedia Computer Science, 87, 32–37. https://doi.org/10.1016/j.procs.2016.05.122

[29] Al-Abdallah, R.Z., & Al-Taani, A.T. (2017). Arabic single-document text summarization using particle swarm
optimization algorithm. Procedia Computer Science, 117, 30–37.
https://doi.org/10.1016/j.procs.2017.10.091

[30] Sarah Aljumah and Lamia Berriche Bi-LSTM-Based Neural Source Code Summarization
https://www.mdpi.com/2076-3417/12/24/12587