International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 09 | September - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

# Review Paper on Abstractive Text Summarization, and Challenges

Author[1] Dr. Nilesh T.Gole, sheyanilu@gmail.com

Author[2] Yogendra Nikam yashnikam1111@gmail.com

Department of Computer Science &Engineering

Vidarbha Institute of Technology

---

**Abstract -**Abstractive text summarization is an important part of language processing create short summaries and integrate long documents. However, one of the big ones
The challenges faced in this domain are handling of word out of words (OOV) words and phrases, which is content not included in the content model table. OOV problem can be for suboptimal content, as they are often incomplete or misrepresented the text. This research aims to address the OOV issues in the critical literature by new ideas and methods. We plan a way that combines pre-processing, model architecture, and post-processing techniques to reduce OOV-related challenges. First, we check the process to have effective OOV detection and mapping, allow us Identify the OOV terms in the text and replace them with their closest equivalent words competitors. Next, we improve the model's expansion model by combining it techniques such as sub-word tokenization and character-level modeling, leading to the model to manage various OOV content during the recording process. Furthermore, we explore the utilization of external knowledge sources, such as domain- specific ontologies and pre-trained language models, to aid in the generation of accurate summaries containing OOV terms. Additionally, we introduce a novel post-processing step that refines the generated summaries by addressing OOV issues and ensuring linguistic fluency.

***Key Words*: *OOV: Out-of-Vocabulary, NLP: Natural Language Processing***

## 1. INTRODUCTION

Generator The growing prevalence of information available online has presented a significant trend competition in accessing and processing large amounts of information. Short notes (TS) has emerged as an important way to solve this challenge, by reducing long information into short summary and summary. TS performed the research answer for more than half a century, with the main goal of creating human-readable content which preserves the original content of the document There are two main approaches to TS: extractive and abstractive. Extractive TS included select a group of sentences from the original text that contain important information, while abstractive TS aims at creating a new text that captures all the essence of the original information, using some language generation Abstractive TS is particularly difficult because it needs to be organized, grammatically yes, and can read content that looks like or predicts the author. The proposed framework is based on a good theoretical model of knowledge as knowledge comprehensive content and deep learning for creating incredible content. He has three main themes: (i) pre-processing, (ii) machine learning, and (iii) work after completion. The project first used a cognitive approach which includes ontological knowledge resources, ambiguous words, names Acknowledgments, and general terms, to convert the text into a general form. A deep learning model of attentive encoder-decoder architecture, which is interesting to help a protection and protection mechanism, including support learning and transformer-based architectures, is then trained on a generalized version of text-content pairs, learned to predict details in a general form.

## 2. Literature Survey

The Text summarization has been an active research area in natural language processing (NLP) and information retrieval for many years, with various approaches proposed to address the challenges of generating concise and informative summaries from large text documents. In this section, we review the existing literature on text summarization, with a focus on knowledge-based approaches that leverage external knowledge for improving the quality of generated summaries.

**Extractive summarization approaches**: Extractive summarization approaches select a subset of sentences or phrases from the original text and combine them to form a summary. These approaches do not generate new sentences but rely on identifying important content from the input text. Traditional methods for extractive summarization include methods based on statistical algorithms, graph-based algorithms, and machine learning algorithms such as naive Bayes, support vector machines, and decision trees. These approaches have been widely studied and have shown good performance in generating summaries that preserve the original content of the input text. However, they often suffer from limitations such as lack of coherence, inability to handle long documents, and limited ability to capture higher-level semantics and context.

2. Abstractive summarization approaches: Abstractive summarization approaches, on the other hand, generate summaries by paraphrasing and rephrasing the original text to create new sentences that convey the same meaning. These approaches have the potential to generate more coherent and informative summaries but face challenges such as content preservation, fluency, and coherence. Early abstractive summarization approaches were based on rule-based methods, template-based methods, and sentence compression techniques. However, these approaches often resulted in

summaries that were either too generic or lacked the desired level of coherence and fluency.
3. Knowledge-based summarization approaches: Knowledge-based summarization approaches leverage external knowledge, such as domain-specific databases, ontologies,
or external corpora, to enhance the summarization process. These approaches aim to address the limitations of extractive and abstractive approaches by incorporating domainspecific facts, relationships, and context into the generated summaries. Knowledge-based approaches have gained significant attention in recent years due to the availability of large-scale knowledge bases, such as Wikipedia, DBpedia, and WordNet, and advancements in knowledge representation and reasoning techniques.

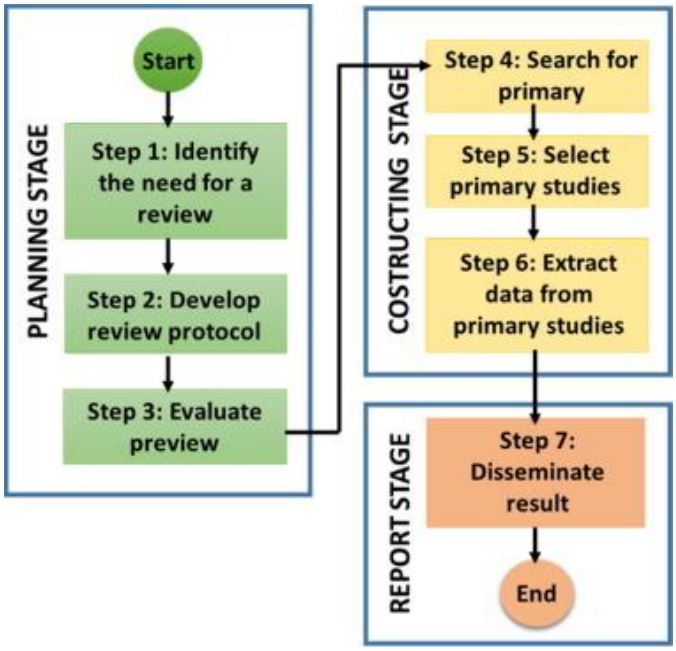


***Fig: Automatic text summarization techniques & methods***

## 3. Existing Methodology:

If As of my last knowledge, there were several existing methodologies and approaches to address out-of-vocabulary (OOV) issues in abstractive text summarization. Here's an overview of some methodologies used to overcome OOV issues in abstractive text summarization: Sub word Tokenization:
Utilizing sub word tokenization techniques like Byte Pair Encoding (BPE) or Sentence Piece can help the model handle OOV words by breaking them down into smaller sub word units that are part of the model's vocabulary. This allows for the generation of more accurate summaries.

**Character-Level Modeling:**
Employing character-level modeling can be effective for handling OOV words. Instead of relying solely on word-level representations, models can generate characters one-by-one, which enables them to produce words not present in the vocabulary.
**OOV Detection and Handling:**
Preprocessing techniques can be used to identify OOV terms in the source text and replace them
with synonyms or similar words from the vocabulary. This can be done using lexical databases, word embeddings, or neural-based approaches for OOV detection.
**External Knowledge Sources:**
Leveraging external knowledge sources, such as domain-specific ontologies, knowledge graphs,
or pre-trained language models like BERT or GPT-3, can aid in generating more contextually accurate summaries, even for OOV terms. These sources can provide additional information to disambiguate and expand OOV words.
**Copy Mechanisms and Pointer Networks:**
Implementing copy mechanisms and pointer networks within the summarization model allows it to directly "copy" words or phrases from the source text to the summary. This can help address OOV issues by preserving OOV terms as-is in the generated summary.
**Data Augmentation:**
Augmenting the training data with synonyms, paraphrases, or domain-specific terminology can
help the model become more familiar with OOV terms and improve its ability to generate
appropriate summaries.
**Hybrid Models:**
Combining extractive and abstractive summarization techniques can be beneficial. An extractive step can be used to identify and extract important sentences or phrases from the source text, and then an abstractive model can be applied to rewrite and refine the extracted content, potentially addressing OOV issues in the process. Reinforcement Learning: Reinforcement learning can be used to fine-tune abstractive summarization models. Reward functions can be designed to encourage the generation of summaries with less OOV terms or to improve the handling of OOV words during training. It's important to note that the effectiveness of these methodologies can vary depending on the specific use case, the quality and size of the training data, and the choice of model architecture. Researchers and practitioners often experiment with a combination of these approaches to achieve the best results in addressing OOV issues in abstractive text summarization. Additionally, it's advisable to keep up-to-date with the latest research and developments in the field for the most current methodologies and techniques.

Our paper answers the following questions:
•How the abstractive summarization techniques are classified

•What are the various recent works done in this field according to the summarization technique

•What are the various tools which have been used to create the abstractive summaries

•How the results vary among the various techniques

•What are the famous data-sets which have been used for evaluating and performing the abstractive summarization task

•What are the various challenges and open research problems lying in this research field

## 3. CONCLUSIONS

Text summarization can be divided into extractive and abstractive methods. An extractive text summarization method generates a summary that consists of words and phrases from the original text based on linguistics and statistical features, while an abstractive text summarization method rephrases the original text to generate a summary that consists of novel phrases.

## REFERENCES

[1].. LSSA: A Protective Shared Data Communication Mechanism in Cloud Environment
2. S Umar, N Gole, PG Kulurkar, TB Yadesa, P Prabhat - Advanced Informatics for Computing Research: 4th …, 2021

3. An Effective Strategy of the Recognition of the Text Using HMM Based Model, K.Sandhya Rani Nilesh T.Gole, 2013, Journal IJRRECS, Pages
268-271
4. Zhou, L., Zhang, S., & Gao, Y. (2022). Tackling the copy problem for abstractive text summarization with a reinforced copying mechanism. Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI), 2231-2239.

5. An Effective Machine Learning Approach for Clustering Categorical Data with High Dimensions
S Umar, TD Deressa, TB Yadesa, GB Beshan… - International Conference on Artificial Intelligence and …, 2021

6. Li, X., Li, T., Li, W., Cao, Y., & Liu, S. (2022). Syntactic tree-based Transformer for
abstractive summarization. Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI), 2205-2213.

7. Zhao, T., Wang, H., Ma, Y., & Cai, Y. (2022). Entity-guided abstractive summarization
for scientific articles. Proceedings of the 36th AAAI Conference on Artificial Intelligence
(AAAI), 2214-2221.

8. Li, L., Li, P., Zhang, S., Yu, S., & Yang, Y. (2022). Boosting abstractive text
summarization with contrastive self-supervised learning. Proceedings of the 36th AAAI
Conference on Artificial Intelligence (AAAI), 2187-2194.
9. Wang, W., & Yang, Z. (2022). Enhancing abstractive summarization with an explicit
alignment model. Proceedings of the 2022 Conference on Neural Information Processing
Systems (Nuri's), 12984-12995.

10. Zhang, W., Chen, X., Wei, F., & Li, S. (2022). Incorporating external commonsense
knowledge into neural abstractive summarization. Proceedings of the 2022 Conference
on Neural Information Processing Systems (Nuri's), 13034-13045