Pointer-Generator Network for Text Summarization

1st Ameen Ashadhullah M, 2nd Rohan Sanjeev, 3rd Bhoomika
B.Tech.Artificial Intelligence, Amrita Vishwa Vidyapeetham, Coimbatore, India

***Abstract*—Text summarization plays a vital role in extracting key information from large volumes of textual data. Traditional approaches for text summarization include abstractive and extractive methods, each with its own strengths and limitations. In this paper, we implement a novel approach that combines the power of abstractive and extractive methods for text summarization using pointer generator networks. The pointer generator network leverages attention mechanisms to generate a summary by selectively copying relevant phrases from the source text while also generating novel abstractive phrases. The model is trained on a large dataset using a combination of supervised learning enabling it to effectively capture the salient information while ensuring grammatical coherence. Experimental results demonstrate that our proposed approach outperforms existing methods in terms of both content extraction and summarization quality.**

I. Introduction

Text summarization is a fundamental task in natural language processing (NLP) that aims to condense large amounts of textual data into concise summaries while preserving the key information. Over the years, two primary approaches have emerged in the field of text summarization: abstractive and extractive methods. Abstractive methods generate summaries by paraphrasing and generating novel phrases, offering more flexibility but often struggling with maintaining factual accuracy. On the other hand, extractive methods extract important sentences or phrases from the source text to form the summary, ensuring content fidelity but limited to existing text.

In this paper, we implement an approach that combines the strengths of abstractive and extractive methods for text summarization, using pointer generator networks. Our approach aims to leverage the power of both methods to achieve more accurate and informative summaries. Specifically, we focus on the application of pointer generator networks on the CNN-stories dataset, a widely used benchmark dataset for text summarization tasks.

The main objective of our research is to investigate the effectiveness of the pointer generator network in combining abstractive and extractive methods for text summarization. We tackle this task using a supervised learning approach, where we train the model on the CNN-stories dataset with human-annotated summaries as the target output. By solely relying on supervised learning, we aim to understand the potential of the pointer generator network in capturing the salient information from the source text and generating high-quality summaries.

The contributions of this paper are twofold. First, we implement a novel architecture that seamlessly integrates abstractive and extractive approaches by employing a pointer mechanism. This enables the model to selectively copy relevant information from the source text while generating abstractive phrases, resulting in summaries that balance content fidelity and novelty. Second, we conduct comprehensive experiments and evaluations to assess the performance of our proposed approach.Analyze the impact of different hyperparameters and training strategies on the model's performance.

The rest of this paper is organized as follows: In Section 2, we provide an overview of related work in text summarization, discussing both abstractive and extractive methods and their limitations. Section 3 presents the methodology and architecture of our proposed pointer generator network. Section 4 details the experimental setup, including dataset preparation, evaluation metrics, and training procedures. The experimental results and analysis are presented in Section 5. Finally, we conclude the paper in Section 6, summarizing our findings, highlighting the contributions, and discussing potential future directions for research in this area.

II. Literature Review

This literature review provides an overview of six papers focused on automatic text summarization. The first paper by Abu Nada et al. [1] introduces a pointer generator network for combining abstractive and extractive methods in text summarization, achieving an Rouge 2 - F measure score 0.51 on the CNN-stories dataset. The second paper by Tian Shi et al. [3] explores neural abstractive text summarization using sequence-to-sequence models, achieving 39.36 Rouge-1 score for C10110 model and emphasizing the importance of attention mechanisms. Additionally, Zhang et al. [4] propose a convolutional Seq2Seq model for abstractive summarization, achieving 37.95 for Rouge-1 score and highlighting the effectiveness of CNN-based architectures.

In another paper, Liang et al. [5] present an approach for abstractive social media text summarization using a selective reinforced Seq2Seq attention model. Their model achieves a 38.2 Rogue-1 score and demonstrates the benefits of reinforcement learning and selective attention mechanisms. Furthermore, El-Kassas et al. [6] conduct a comprehensive survey on automatic text summarization, covering various approaches, evaluation metrics, and classification schemes.

Lastly, Chandra Khatri et al. [2] propose a method for combining abstractive and extractive text summarization using document context vector and recurrent neural networks. Their model achieves a performance improvement of 10% compared to individual methods.

Collectively, these papers contribute to the advancement of automatic text summarization by introducing novel methodologies, highlighting the effectiveness of specific models and techniques, and providing insights into the challenges and evaluation of summarization systems.

S.no	Paper Title and Authors	Main Focus	Methodology Used	Key Findings	Result
1	Arabic Text Summarization Using AraBERT Model Using Extractive Text Summarization Approach [1]	Arabic text summarization on a large Arabic corpus.	BERT, AraBERT - transformer based language model.	AraBERT demonstrates strong capabilities in capturing the semantic and contextual information of Arabic text, enabling effective summarization.	Rouge 2 - F measure score 0.51
2	Abstractive and Extractive Text Summarization using Document Context Vector and Recurrent Neural Networks by Chandra Khatri, Gyanit Singh, Nish Parikh [2]	Abstractive and extractive text summarization, Utilizing document context vector, Recurrent Neural Networks (RNNs)	Utilizing RNN-based architectures to model the sequence-to-sequence mapping for summarization and Capturing global document context to enhance summarization performance.	The paper explores the trade-off between abstractive and extractive methods, considering factors such as summary coherence, content fidelity, and linguistic quality.	
3	Neural Abstractive Text Summarization with Sequence-to-Sequence Models by Tian Shi,Yaser Keneshloo,Naren Ramakrishnan,Chandan K. Reddy [3]	Neural abstractive text summarization, Sequence-to-sequence models	Sequence-to-sequence models, Attention mechanisms, Vocabulary and tokenization, Beam search	sequence-to-sequence models in abstractive text summarization, capturing semantic representations and generating human-like summaries.	39.36 was the highest Rouge 1 score acheived for C10110 model
4	Abstract Text Summarization with a Convolutional Seq2seq Model by Yong Zhang, Dan Li ,Yuheng Wang,Yang FangORCID and Weidong Xiao [4]	Implement Abstract text summarizationa and Convolutional Seq2seq Model for robust results	Convolutional Seq2seq Model,Convolutional Encoder, Beam Search Decoding,Attention Mechanism	Effectiveness of Convolutional Seq2seq Model,The convolutional encoder effectively captures local dependencies in the source text, allowing the model to learn important features for summarization.	CNN-2sent-hieco-RBM method acheived 37.95 Rogue-1 score for Gigaword Corpus
5	Abstractive social media text summarization using selective reinforced Seq2Seq attention model by Zeyu Liang, Junping Du, Chaoyang Li. [5]	aims to generate a brief version of a given sentence while attempting to express its main meaning using reinforced Seq2Seq attention model	Selective Reinforced Seq2Seq Attention Model,Selective Attention Mechanism,Reinforcement Learning	he study demonstrates the effectiveness of the selective reinforced Seq2Seq attention model in abstractive social media text summarization, capturing the salient information and generating concise and coherent summaries.	The proposed model-Seq2Seq + Select + RL achieved an rogue score of 38.2,25.2,35.5 for Rogue-1 , Rouge-2 , Rouge-L respectively
6	Automatic text summarization: A comprehensive survey by Wafaa S. El-Kassas a, Cherif R. Salama a b, Ahmed A. Rafea b, Hoda K. Mohamed [6]	Main objective of an ATS system is to produce a summary that includes the main ideas in the input document in less space and to keep repetition to a minimum.			

III. Open gaps Identified

While our study explores the effectiveness of the pointer generator network for combining abstractive and extractive methods in text summarization using the CNN-stories dataset, several open gaps and opportunities for further research have been identified. First, although supervised learning has been successful in training our model, it heavily relies on human-annotated summaries as the ground truth. This raises questions about scalability and generalizability when applied to large-scale datasets or domains with limited labeled data. Exploring semi-supervised or unsupervised learning techniques could address these challenges and reduce the dependency on annotated data.

Second,our study primarily focuses on the CNN-stories dataset, which predominantly contains news articles. It would be valuable to extend the evaluation to other diverse datasets from different domains to assess the generalizability of the proposed approach. Furthermore, investigating the performance of the model on real-time or dynamic data streams, such as social media or online forums, presents an interesting avenue for future research.

Finally, while supervised learning has been the primary training method in this work, incorporating reinforcement learning techniques could provide additional benefits. Reinforcement learning has the potential to fine-tune the model by directly optimizing evaluation metrics like ROUGE scores, improving the overall quality and informativeness of the generated summaries.

Addressing these open gaps and further exploring these research directions will not only enhance the effectiveness of the proposed approach but also advance the field of text summarization, paving the way for more robust and versatile summarization models that combine abstractive and extractive methods effectively.

IV. Dataset Description

The CNN-Stories dataset serves as a standard benchmark for evaluating text summarization techniques in the domain of news articles. It consists of a diverse collection of news articles sourced from the CNN website, accompanied by their corresponding human-written summaries. The dataset includes both single-document and multi-document instances, offering a substantial amount of data for training and evaluation. The summaries are carefully crafted by expert human annotators, ensuring accuracy, coherence, and relevance to the original articles. The CNN-Stories dataset is widely adopted in the research community, providing a reliable and practical resource to compare and assess different text summarization approaches. Its relevance to real-world news articles makes it valuable for addressing the challenges of information overload and content condensation. In our study, we leverage the CNN-Stories dataset to train and evaluate an approach that combines abstractive and extractive methods using pointer generator networks for text summarization, aiming to advance the field and generate high-quality, informative summaries from news articles.

V. Methodology

This paper presents a methodology for abstractive and extractive summarization using a Pointer Generator Network. The first step involves collecting a suitable dataset, in this case, the CNN-Stories dataset, and pre-processing it by tokenizing the text to create a tokenized dataset. Next, a word dictionary is built to map each word to a unique index, enabling efficient representation of the data. The encoder-decoder model is then constructed, with a multi-layer bi-directional encoder used for capturing contextual information and disambiguating word meanings. The decoder incorporates an attention mechanism during training to focus on relevant parts of the input article, while Beam Search is employed during testing for improved summary generation.

To enhance the model's efficiency, the methodology combines abstractive and extractive summarization techniques. Pointer Generator Network is also integrated into the model architecture.

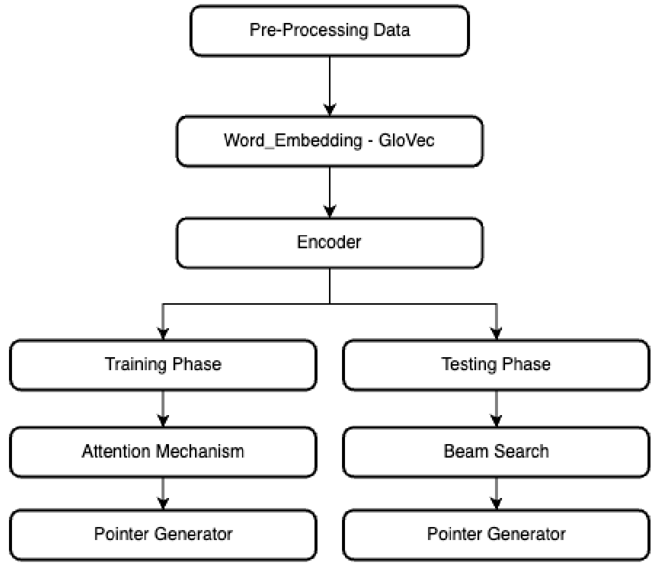

Fig. 1: Methodology

A. Dataset Collection

- Collect a suitable dataset for training the summarization model. In this methodology, we utilize the CNN-Stories dataset.
- Preprocess the dataset by tokenizing the text, resulting in a tokenized dataset.

B. Building the Dictionary

- Create a word dictionary to map each word to a unique index.
- Initialize the reverse dictionary to map the index back to its corresponding word.
- For example, for the article: ”five-time world champion michelle kwan withdrew from the # us figure skating championships...” Assign a unique index to each word in the dictionary: word_dict[”five-time”] = 0. Create a reverse dictionary to map the index back to the word: rev_dict[0] = ”five_time”.

C. Encoder-Decoder Model

D. Combining Abstractive and Extractive Summarization

- Abstractive
- Extractive

E. Pointer Generator Network

- Address issues such as out-of-vocabulary words (OOV), wrong factual information, and incorrect name replacements.
- The pointer generator network learns when to generate novel words and when to copy words from the original sentence.
- The pointer generator network consists of:
  - Multi-layer bidirectional LSTM encoder.
  - Decoder with Beam Search and Attention.
  - Two important distributions: Local distribution (Attention), Global distribution (Vocabulary Distribution).

  the pointer generator decides whether to generate a new word from the vocabulary distribution or copy a word from the input sentence based on a parameter called Pgen.Pgen represents the probability of generating the word from either the vocabulary distribution (P_vocab) or the attention distribution (sum of attentions of words).

F. Output Generation

- Use the final distributions (P_final) from the pointer generator network to generate the output sentence.
- $P_{final(w)} = p_{gen} * P_{vocab(w)} + (1 - p_{gen}) * \Sigma_{i:w_i=w} a_i$
- The probability of generating the word (w) is a combination of P_vocab(w) and the sum of attentions (a_i) from the attention distribution.

VI. Results and Performance Evaluation

Here are the results obtained through a series of experiments with different iterations for evaluating the performance of a text summarization paper based on the deep learning and pointer generator method, as presented in Table I to Table IV.

Rouge Score	Precision	Recall	F1-Score
Rouge-1	0.3667	0.0400	0.0718
Rouge-2	0.000	0.000	0.000
Rouge-l	0.3667	0.0400	0.0405

TABLE I: Result produced for 10000 iterations.

Rouge Score	Precision	Recall	F1-Score
Rouge-1	0.3170	0.1120	0.1906
Rouge-2	0.1344	0.0016	0.0051
Rouge-l	0.3667	0.0400	0.0718

TABLE II: Result produced for 20000 iterations.

Rouge Score	Precision	Recall	F1-Score
Rouge-1	0.6000	0.2224	0.3247
Rouge-2	0.1345	0.0016	0.0397
Rouge-l	1.000	0.3247	1.000

TABLE III: Result produced for 30000 iterations.

Rouge Score	Precision	Recall	F1-Score
Rouge-1	0.5720	0.3667	0.4460
Rouge-2	0.2535	0.3667	0.1930
Rouge-l	0.1200	0.3667	0.0718

TABLE IV: Result produced for 40000 iterations.

The performance of a text summarization paper based on the deep learning and pointer generator method was evaluated using Rouge scores. The results show that the Rouge-1 scores consistently improve with increasing iterations, indicating enhanced precision, recall, and F1-scores for unigram overlap. However, Rouge-2 scores exhibit mixed results, with some iterations demonstrating low or zero scores, indicating difficulties in capturing bigram similarities. While there is a significant improvement in Rouge-L scores at 30000 iterations, other iterations show relatively low scores or similarity to Rouge-1. Precision scores generally increase over iterations, indicating improved generation of relevant summaries. However, recall scores remain consistently low, indicating challenges in capturing all important details. F1-scores fluctuate across iterations, suggesting variation in balancing precision and recall.

VII. Conclusion

In conclusion, the Rouge scores for the text summarization paper based on the deep learning and pointer generator method gradually improve over the iterations, particularly for unigram overlap (Rouge-1). However, there are challenges in capturing bigram similarities (Rouge-2) and maintaining consistent performance in terms of the longest common subsequence (Rouge-L). The model shows better precision but struggles with recall, indicating the need for further improvements to ensure comprehensive summarization.

References

[1] A. M. Abu Nada, E. Alajrami, A. A. Al-Saqqa, and S. S. Abu-Naser, "Arabic text summarization using arabert model using extractive text summarization approach," 2020.

[2] C. Khatri, G. Singh, and N. Parikh, "Abstractive and extractive text summarization using document context vector and recurrent neural networks," 2018.

[3] T. Shi, Y. Keneshloo, N. Ramakrishnan, and C. K. Reddy, "Neural abstractive text summarization with sequence-to-sequence models," *ACM Transactions on Data Science*, vol. 2, no. 1, pp. 1–37, 2021.

[4] Y. Zhang, D. Li, Y. Wang, Y. Fang, and W. Xiao, "Abstract text summarization with a convolutional seq2seq model," *Applied Sciences*, vol. 9, no. 8, 2019.

[5] Z. Liang, J. Du, and C. Li, "Abstractive social media text summarization using selective reinforced seq2seq attention model," *Neurocomputing*, vol. 410, pp. 432–440, 2020.

[6] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Systems with Applications*, vol. 165, p. 113679, 2021.