International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6454
Article Received: 09 February 2023 Revised: 13 March 2023 Accepted: 22 March 2023

Abstractive Summarization with Efficient Transformer Based Approach

Madhuri P. Karnik[1], Dr. D.V. Kodavade[2]
[1]Ph.D Scholar, Shivaji University, Kolhapur, Maharashtra, India
madhuri.chavan@viit.ac.in
[2]Professor, D.K.T.E. Society's Textile & Engineering, Institute, Ichalkaranji, Maharashtra, India
dvkodavade@gmail.com

Abstract— One of the most significant research areas is how to make a document smaller while keeping its essential information because of the rapid proliferation of online data. This information must be summarized in order to recover meaningful knowledge in an acceptable time. Text summarization is what it's called. Extractive and abstractive text summarization are the two types of summarization. In current years, the arena of abstractive text summarization has become increasingly popular. Abstractive Text Summarization (ATS) aims to extract the most vital content from a text corpus and condense it into a shorter text while maintaining its meaning and semantic and grammatical accuracy. Deep learning architectures have entered a new phase in natural language processing (NLP). Many studies have demonstrated the competitive performance of innovative architectures including recurrent neural network (RNN), Attention Mechanism and LSTM among others. Transformer, a recently presented model, relies on the attention process. In this paper, abstractive text summarization is accomplished using a basic Transformer model, a Transformer with a pointer generation network (PGN) and coverage mechanism, a Fastformer architecture and Fastformer with pointer generation network (PGN) and coverage mechanism. We compare these architectures after careful and thorough hyperparameter adjustment. In the experiment the standard CNN/DM dataset is used to test these architectures on the job of abstractive summarization.

Keywords: - Summarization, Transformer, Pointer Generator Network, Coverage, Self-Attention, Fastformer.

I. INTRODUCTION

The number of textual data, such as articles, news, documents, reviews, etc., has dramatically expanded in the modern era, becoming a useful resource for information extraction and analysis. This knowledge must be summarized in order to recover it in an acceptable amount of time. The two main methods of summarization are extraction and abstraction. The first attempts in summarization centered on extractive methods, which identify words or phrases in a document that best convey its key points. While second summarizing attempts focused on generating short summary for the original document.

Due to the difficulty of NLP abstractive text summarizations became a challenging task hence many of the researchers have shifted their attention towards abstractive text summarization. The extraction approach only pulls sentences as a summary from the source material; however, the abstract methodology might generate new words and phrases for the original content. It is thought to be a difficult task for a machine to perform despite the fact that ATS rephrases the original text to construct new concepts that might not be in the original text. When people are having trouble locating the pertinent information within a document or documents, ATS can be quite beneficial. Another objective of ATS is a review that condenses the essential ideas from the input content into a concise amount of text. As a result, the user will gain from the summaries that were automatically prepared, saving them a lot of time and work.

Because ATS necessitates a thorough knowledge of the source in order to produce the summary, it requires real-word knowledge as well as substantial NLP. Abstractive summarization is more like what a person would do in real life. He imagines the text, compares it to his recollection and relevant knowledge, and then re-creates the text's core in a short sentence. As a result, abstractive summarizing is more difficult than extractive summarization because the model must break down the source corpus to the token level and recreate the target sentences. Different structured and semantic-based approaches were used previously for abstractive text summarization.

In the publication [1], a survey and observations of extractive and abstractive text summarization methods are described, and in the paper [2], a survey on several approaches to automatic abstractive text summarization is presented. In 2015 [3], abstractive text summarization used deep learning techniques for the first time, and the proposed model utilized an encoder-decoder design. Several deep learning models have emerged in recent years, including RNN, CNN, seq2seq, LSTM models, have been widely used for this purpose and have produced great results.

Although RNN has wide adaptability, there are some issues with RNN. The gradient expanding and vanishing issues with RNNs make it difficult to learn from lengthy data sequences. To retain the inputs for longer period of time, LSTM networks are used, however, it might not be enough to resolve the issue entirely. The recent technique applied in abstractive summarization, such as Transformer is introduced in [11]. It was totally based on attention mechanisms and substitutes multi-headed self-attention for the recurrent layers most frequently found in encoder-decoder architectures [11].

The section 2 provides an overview of some of the deep learning models and approaches utilized for abstractive text summarization. The different models are described in section 3 such as Transformer, Pointer Generator Network, and Fastformer. Experimental results including datasets, preprocessing methods, evaluation metrics, result analysis, test findings are discussed in section 4 and the study's conclusion and next steps are presented in section 5.

II. RELATED WORK

Because of the intricacy of NLP, abstractive text summarization has become a tough task, and many researchers have shifted their focus to this area. Structured and semantic-based approaches to abstractive text summarization are examples. Deep learning is used in a variety of NLP tasks since it allows for the learning of multilevel hierarchical data representations. RNNs, CNNs, and seq2seq models are among the deep learning models that have been hired for abstractive summarization.

In 2015, deep learning techniques were used for the first time in abstractive text summarization [3], with an encoder-decoder architecture as the proposed model. Deep learning approaches have shown to be effective in many applications, and they have been widely used in recent years. Some current research in the ATS area is mentioned here.

A data-driven technique to abstractive phrase summarization was developed by A. M. Rush et al. [3]. For each word of the summary trained on the input text, their technique used a local attention-based model. They have concentrated on the task of summarizing sentences. The creation of paragraph-level summaries is considered a step forward in their work.

A conditional RNN that generates an input sentence summary was given by Chopra, S., et al. [4]. A unique convolutional attention-based encoder provides the conditioning, ensuring that the decoder centers around the correct input words at each step of the generating process. The proposed model is an expansion of the model for a similar issue by Rush et al. (2015). Chopra S. et al. [4] used a feed-forward neural language model instead of an RNN for creating summaries.

Nallapati et al. [5] coupled the attention model with the RNN in order to address problems with keyword modelling and capture the hierarchical structure between the sentence and word. They used the bidirectional encoder with GRU-RNN and unidirectional decoder with GRU-RNN. On the hidden states of the source, they employed the attention model, and on the target vocabulary to generate words, they used the softmax layer. Junyang Lin et al.[6], presented global encoder model for abstractive summarization. The model relies on the seq2seq model with attention. For the encoder, a convolutional gated unit is used. In context of the results from the RNN encoder, by using CNN the global encoding improves the representation of the source context which further develops the word relationship with global context. The performance assessment is done using GPU based approach with Pytorch.

The deep learning method using LSTM-CNN based framework is presented for abstractive text summarization by Shengli Song et al. [7]. The new ATSDL model extracts essential phrases from the source text using a phrase extraction method known as MOSP (Multiple Order Semantic Parsing), and then learns phrase association. The phrase sequence that the model generates after training will meet the requirements for syntactic structure.

To obtain relation amongst other words thereby ignoring the grammatical approach of identifying verbs, pronouns etc., the phrases are considered either as subjective or as objective. The convolutional neural network-based encoder is used with LSTM model to perform quicker processing in forward training method. ROUGE parameter is utilized to assess the method along with average count of different phrases is additionally thought of.

Li et al., [8] utilized the seq2seq encoder-decoder model to produce the abstractive summaries. To increase the quality of summaries, they took into account the text's latent structural information. To construct the summary, they translated the source code into hidden states using the deep recurrent generative decoder before translating it back to the original word sequences.

A query-based strategy for abstractive summarization was presented by Baumel et al.[9]. The query focused summarization used is based on relevance model and generic summarization for obtaining efficient summary along with redundancy removal. For getting an efficient summary and removing redundancy, the query focused summarization utilized is based on relevance model and generic summarization. In the seq2seq model, a pre-calculated relevance score is assigned to each word to evaluate the weight. Because Softmax employs the attention model to normalize the weight, therefore the rescaling is done with the cosine transform, which enhances the ROUGE outcomes. The performance is also evaluated using TF-IDF and word count metrics.

The authors [10] employed a bi-directional RNN with LSTMs in the encoding layer and an attention model in the decoding layer. Their main objective was to develop a more

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6454
***Article Received**: 09 February 2023 **Revised**: 13 March 2023 **Accepted**: 22 March 2023*

effective abstractive text summarizer by increasing the effectiveness and decreasing the training loss of the seq2seq model. They were successful in reducing the training loss to 0.036, and their abstractive text summarizer produces a short summary of English-to-English text.

Transformer is a relatively new technique applied in abstractive summarization. In this section of the study, the research employing Transformer for ATS application is discussed. Vaswani et al. [11] proposed the Transformer, a model architecture that foregoes recurrence in favor of an attention mechanism to draw global interdependence between input and output. A lot more parallelization is possible with the Transformer. Authors have developed the Transformer, which is based purely on attention, in which multi-headed self-attention replaces the recurrent layers that are usually employed in encoder-decoder designs.

Abigail et al. [12] of Stanford University developed a generative pointer network and coverage technique to overcome the difficulties of semantically inaccurate and repetitive abstractive text summaries in 2017. A hybrid PGN that they developed can replicate words from a source text by pointing at them, allowing for more accurate information reproduction while still allowing the generator to generate new words. Second, coverage is used to keep track of what has been summarized, which reduces the need for repeating.

In order to include the multi-head self-attention mechanism into the fundamental encoder-decoder model and semantically improve the structure of abstractive text summaries, Qian guo et al. [13] proposed an MS-Pointer Network in 2019. A pointer network is developed on the seq2seq using a multi-head attention strategy to address the issue of out of vocabulary (OOV).

LTABS model was introduced by Xin Liu et al.[14]. The transformer serves as the LSTM attention matrix in this model, which uses the LSTM hidden layer result as location information to enhance local attention. They have thought of using a copy mechanism to address the issue of problem of OOV.

A combined summary generation model based on improved transformer has been proposed by Xin Liu and Liutong Xu [15]. By using a RNN, this model can focus attention on a single sentence while also providing sequence information to the periodical transformer model. The Transformer model is used in the generation stage to comprehend the long-distance dependence between words, and the consistency loss function and pointer mechanism make the summary statement more consistent with the original meaning. They presented the TP-EABS (Transformer added Pointer and combine the Extractive and Abstractive techniques) model, which incorporates the benefits of two types of algorithms as well as the transformer model. By incorporating a pointer mechanism into the transformer model, words from the original text can be copied.

The text summarization problem was investigated using Sequence-to-sequence RNNs and Transfer Learning with a Unified Text to-Text Transformer approaches in this research presented by Ekaterina Zolotareva, et al. [16]. In order to lessen duplication, factual errors, and OOV mishandling in abstractive summarization, Jon Deaton et al.[17] have presented new transformer models that use PGN, coverage, and n-gram blocking. They have demonstrated that each of these models outperforms the functionality of the earlier models.

To create abstracts, Wen Kai and Zhou Lingyu [18] suggested a pre-training strategy based on BERT and LeakGAN model. The BERT model is paired with the LeakGAN model in this paper to provide a summary. The BERT model is pre-trained on the higher-quality word vector, which is then fed into the LeakGAN model for abstract creation. The LeakGAN model is also somewhat tuned at the same time. The discriminator in the LeakGAN model includes an attention mechanism, which directs the discriminator's attention to important information connected to input data and feedback information, hence enhancing feature output quality.

Four new ATS models with a seq2seq structure were proposed by Jianwen J et al. [19] using an attention-based bidirectional LSTM. Additionally, these models contain improvements for tackling the issue of out-of-vocabulary terms, suppressing repeated words, and minimizing the growth of cumulative errors in generated text.

Jiaming Sun et al. [20] introduced the Summarization Transformer with Template-aware Representation (STTR), a new abstractive summarization paradigm. which employs a document representation shifting loss and a template-aware document encoding module to maintain important information while filtering template noise.

The work that the researchers did for the ATS application employing Transformer and Pointer Generator Network is described in the following section of the study. In order to detect false news, as a feature extractor for a task, Soheil Esmaeilzadeh et al. [21] employed their text summarizing model that summarizes news articles before categorizing them. The outcomes were contrasted with the classification made using only the news article's original text.

Transformer LM was used by Urvashi Khandelwal et al.[22], to encodes the source and produce the summary using a pre-trained decoder-only network. This ensures that before the fine-tuning process, all network parameters, including those governing attention over source states, have been pre-trained. In limited-data scenarios, their pre-trained Transformer LM outperforms pre-trained Transformer encoder-decoder networks, according to their experiments.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6454
Article Received: 09 February 2023 Revised: 13 March 2023 Accepted: 22 March 2023

Vrinda Vasavada et al.[23] examined See et al.'s pointer-generator model[10], then extended the model by experimenting with transformers [11] instead of typical LSTM encoders and decoders to explore potential areas of improvement.

DANG Trung Anh et al.[24] improved the network by using the two most recent word embedding techniques, Word2vec and Fasttext, to more properly capture the semantic content of words. Some OOV terms that were previously identified as unknown tokens can now be properly embedded and evaluated in summary generation.

A seq2seq fine-tuning toolset called s2s-ft, developed by Hangbo Bao et al. [25], makes use of pretrained Transformers for conditional generation tasks. They have implemented seq2seq fine-tuning algorithms. Through neural abstractive summarization, Sandeep Subramanian et al. [26] have described a method for producing abstractive summaries of large materials with over a thousand words. Before generating a summary, they performed a simple extraction phase, which was then utilized to provide relevant information to the transformer language model before it was requested to provide a summary. They have shown that using an extractive and abstractive process, Transformer language models are capable of creating high-quality summaries of large text sequences.

RC-Transformer (RCT) is a new abstractive summarization model presented by Tian Cai et al [27]. Not only does the model learn long-term dependencies, but it also overcomes Transformer's disregard for word order data. To capture the sequential context representations, they have added an extra RNN-based encoder to the Transformer. Also, they developed a convolution module to effectively extract vital information by filtering the sequential context with local importance. Transformers also succeed in abstractive summarization tasks, according to Nima Sanjabi [28].

For learning the conditional probability of an output sequence using discrete tokens that match to locations in an input sequence, Oriol V. et al. [29] suggested a new neural architecture. Using a recently suggested neural attention mechanism, their model overcomes the problem of changing size output dictionaries. It differs from earlier attention attempts in that it chooses an input sequence member as the output using attention as a pointer rather than combining an encoder's hidden units with a context vector at each decoder step. This structure is known as a Pointer Net (Ptr-Net). With input attention, Ptr-Nets outperform seq2seq and generalize to output dictionaries of different sizes.

The authors [30] most recently used the pre-trained BERT model for book summaries. They chose the phrases from the manuscript exactly as they are, which helps to preserve the book's key terminology and keeps the book's phrasing intact. The proposed work is tested in the paper [31] for book, in particular as a long document for evaluating the model's efficacy. Based on the findings, the author has determined that BERT uses the best wording possible to distil a lot of information into a concise summary.

III. MODEL

In this section, four models are presented 1) Transformer model [11] (2) Transformer with PGN and coverage model [17] 3) Fastformer model [28] and 4) Fastformer with PGN and coverage model.

A. *Transformer Model*

Google researchers [11] proposed the transformer, a novel network architecture. It was entirely dependent on attention mechanism. The first sequence transduction model is the transformer that can establish the illustrations of the input and output without the need of convolution or sequence recurrence. The architecture of a transformer model is depicted in Fig.1 [11]. The different normalization and multi-head attention layers, stacked self-attention layers, and point-wise, layers for the encoder and decoder are entirely coupled and all are depicted in Fig. 1.

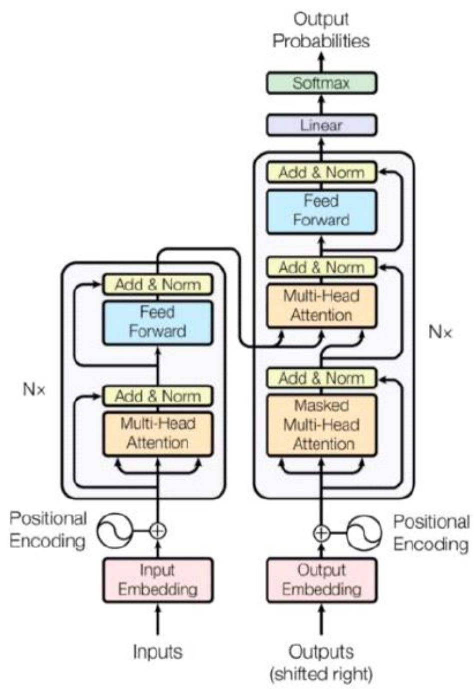

Fig. 1: The Transformer - model architecture [11].

The Transformer model's sole support is made up of Multihead attention layers [11]. To recall the input word sequence, it uses positional encoding and attention layers. The encoder and decoder layers, as well as a multi-head attention layer and feed forward network layers, are connected in the transformer model

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6454
***Article Received**: 09 February 2023 **Revised**: 13 March 2023 **Accepted**: 22 March 2023*

[11]. The model recalls the location and order of words, leading to positional encoding, with the aid of the sine and cosine functions. Self-attention is a method used by the multi-head attention layer of the encoder and decoder layer [11]. In order to create query (Q), key (K), and value (V) vectors, the input is routed through three linked layers [11]. They simultaneously estimated the attention function for the set of queries contained in the Q matrix. The matrices K and V also include the keys and values combined, and the following formula is used to calculate the output matrix:

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad [11]$$

where d_k is the dimension of the input K. Then, to create multi-head attention, self-attention is applied to n different vectors [11].

B. Transformer And Pointer Generator Network Model

Oriol Vinyals et al. [29] were the first to propose the pointer-generator. Since then, the pointer-generator has been used in a variety of NLP. The authors' pointer-generator network [17] is based on See et al. [12], who used their model to train an RNN and achieved cutting-edge results in abstractive summarization. The hybrid PGN in the See et al. [12] model can point at words in the source text and copy them, providing a more exact replication of the information, while yet keeping the capacity to use the generator to create new words. Second, by utilizing coverage to keep track of what was summarized, they reduced repetition. On seq2seq tasks such as machine translation, transformers outperformed Recurrent Neural Networks. Summaries produced by transformer models are repetitious and usually factually inaccurate.

Using PGN, coverage, and n-gram blocking, authors [17] improved the transformer model to reduce repetition, factual errors, and out of vocabulary (OOV) mishandling in abstractive summarization. Following Fig.2 represents the Pointer-generator network on a transformer proposed by authors [17].

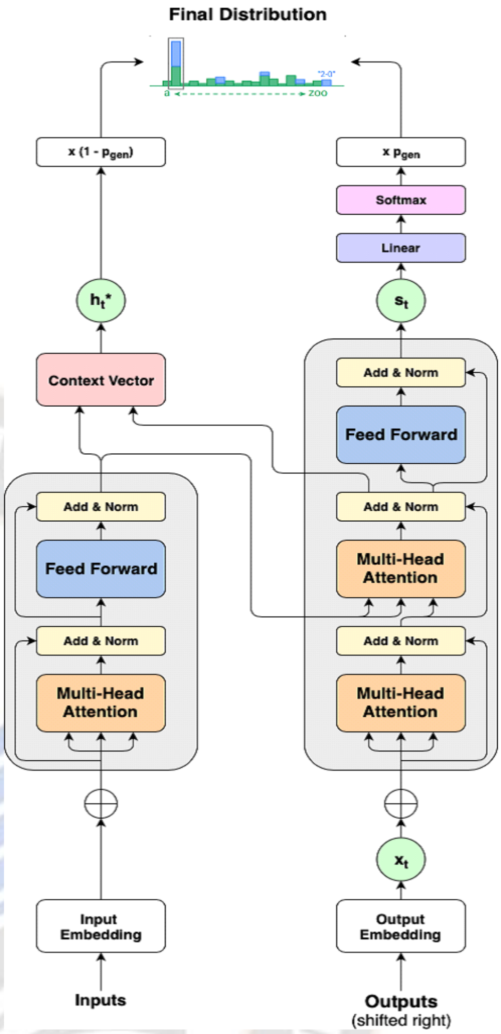

Fig. 2 : Pointer-generator network on a transformer [17]

C. Fastformer Model

For text summarization Transformer is a widely used model. The quadratic complexity of the input sequence length, however, makes it inefficient. Despite the fact that there are numerous methods for accelerating Transformers, they are either ineffective or inefficient for extended sequences. An effective Transformer version based on additive attention that can accomplish context modelling in linear complexity has been developed by Chuhan Wu et al. [32] called Fastformer. This model handles long sequences efficiently. In Fastformer model, the input attention query matrix was initially condensed via an additive attention method into a global query vector. Then, they use element-wise product to describe the interaction between the attention key and the global query vector to create a global context-aware key matrix, which is then further condensed into a global key vector using additive attention. The global key and attention value are then combined using element-wise product, after which a linear transformation is used to calculate the global context-aware attention value. The final result is created by adding the initial attention query and the global context-

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6454
***Article Received**: 09 February 2023 **Revised**: 13 March 2023 **Accepted**: 22 March 2023*

aware attention value. This allows the computational complexity to be lowered to linearity while still efficiently capturing the contextual data from the input sequence. Fig. 3 depicts architecture of the Fastformer model.

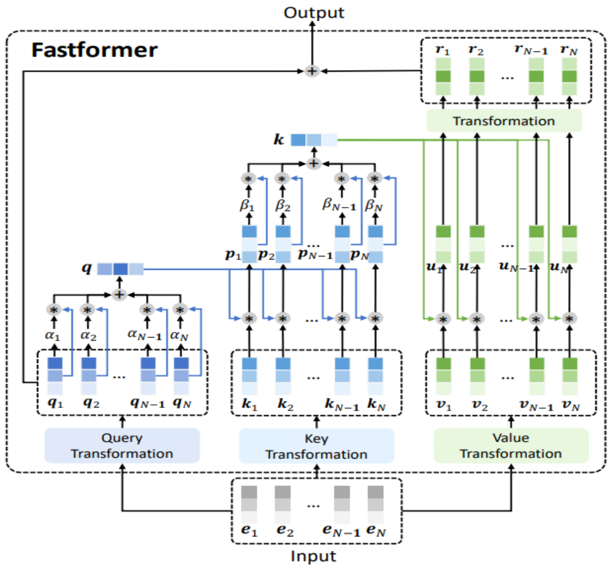

Fig. 3 : Fastformer model [32]

D. Fastformer with Pointer generator network and coverage

This is an approach where we combine Fastformer model with Pointer generator network and coverage. As mentioned in the section above, pointer-generator networks, coverage, and the transformer model all work together to improve abstractive summarization by reducing repetition, factual errors, and out of vocabulary mishandling [17]. When combined pointer generator network and coverage with the Fastformer model, the results are further enhanced by linear complexity.

IV. EXPERIMENTAL DETAILS

This section describes the dataset used for this application as well as the implementation details for various approaches.

i) Dataset

The models are assessed using the reasonably big CNN/DM dataset. It has more than 200 million words in it. CNN/DM dataset (corpus) is a collection of articles. News stories and highlight sentences make up the information. Each article has a few highlighted elements that together make up the overall overview. This dataset was created to aid in the development of models that can summarize long paragraphs of text in one or two sentences. The Stanford coreNLP Java toolkit, which is available online, is used to preprocess and tokenize the data.

ii) Preprocessing

In many applications, data preparation is a critical and important step. The CNN/DM dataset includes long news stories as well as brief summaries for comparison. The raw dataset was cleaned using a variety of pre-processing techniques such as eliminating whitespace, duplication, stopwords, lower casing, punctuation removal, stemming, and so on.

iii) Evaluation Methods

We use ROUGE score, a standard scoring system for summarization tasks, to assess the generated summaries. Rouge is a statistical indicator that compares the frequency of tokens in the source and destination languages. This metric is frequently used by researchers to determine the level of consistency between the system generated summary and the reference summary by counting the number of words that match. To assess the quality of a summary, ROUGE counts overlapping units like the n-gram, word sequences, and word pairings. The overlap of n-grams between the system generated summary and the ideal reference summary is denoted by ROUGE-N. For example, ROUGE-1(R-1), ROUGE-2(R-2), ROUGE-3(R-3) and ROUGE-L(R-L) (Longest Common Subsequence).

iv) Implementation Details

TensorFlow 2.6.0 and Python 3.9.7 are the deep learning framework and experimental programming language used in the model, respectively. The NVIDIA GeForce RTX 3050 GPU server with 16 GB of RAM is the hardware environment used.

v) Results and Analysis

In this study, we examine how well the summarization models such as (1) Transformer, (2) Transformer with PGN and Coverage (3) Fastformer model and (4) Fastformer with PGN and Coverage performed. Table 1 summarizes the major findings. The F1 scores for R-1, R-2, R-3, R-L are displayed between the reference summary and the evaluation summary. After contrasting the performance of several models with those of numerous current techniques, results are shown on the CNN/DM dataset.

TABLE 1. ROUGE scores of different text summarization models on CNN/DM dataset.

Models	**R-1**	**R-2**	**R-3**	**R-L**
Transformer	36.40	24.34	20.09	41.57
Fastformer	38.68	27.89	23.42	42.57
Transformer+ PGN + Coverage	41.74	29.32	23.77	46.10
Fastformer+ PGN + Coverage	**47.22**	**35.87**	**31.17**	**51.06**

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4
DOI: https://doi.org/10.17762/ijritcc.v11i4.6454
***Article Received**: 09 February 2023 **Revised**: 13 March 2023 **Accepted**: 22 March 2023*

The Fig.4 shows comparative performance analysis which is done between aforementioned models.

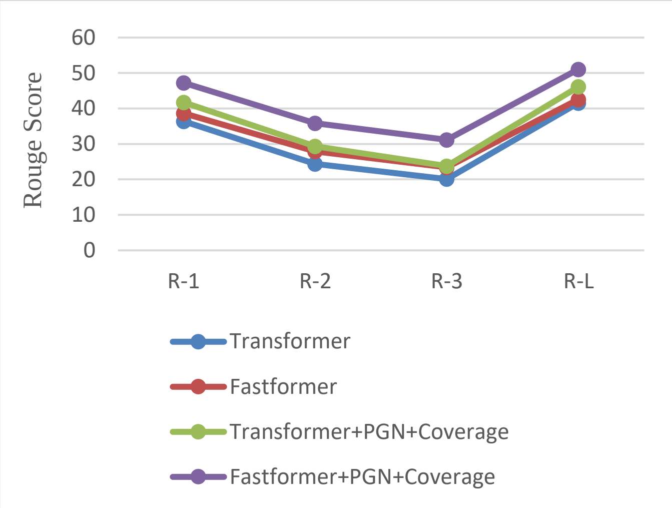

Fig. 4: Comparative Analysis of F1 scores for all 4 models.

Additionally, we give an example to show how models produce superior summaries to one another. The example is shown in Table 2.

TABLE 2 Output of 4 ATS models on a News article

Original Article (truncated): marseille , france cnn the french prosecutor leading an investigation into the crash of germanwings flight 9525 insisted wednesday that he was not aware of any video footage from on board the plane . marseille prosecutor brice robin told cnn that `` so far no videos were used in the crash investigation . '' he added , `` a person who has such a video needs to immediately give it to the investigators . '' (…)
Reference Summary: marseille prosecutor says `` so far no videos were used in the crash investigation '' despite media reports . journalists at bild and paris match are `` very confident '' the video clip is real , an editor says . andreas lubitz had informed his lufthansa training school of an episode of severe depression , airline says .
Transformer-Summary: marseille prosecutor leading an investigation into the crash of germanwings flight 9525 insisted wednesday that he was not aware of any video footage from on board the plane `` so far no videos were used in the crash investigation andreas of an episode of severe depression , airline says .
FastFormer - Summary: marseille leading an investigation into the crash of germanwings flight 9525 insisted wednesday that he was not aware of any video footage from on board the plane so far no videos were used in the crash investigation andreas an episode of severe depression , airline says .
Transformer + Pointer generator network+Coverage Summary: marseille leading an investigation into the crash of germanwings flight 9525 insisted wednesday that he was not aware of any video footage from on board the plane so far no videos were used in the crash investigation andreas an episode of severe depression , airline says .
Fastformer + Pointer generator network + Coverage Summary: marseille france cnn the french prosecutor leading an investigation into the crash of germanwings flight 9525 insisted wednesday that he was not aware of any video footage from on board the plane prosecutor brice robin told cnn that `` so far no videos were used in the crash investigation andreas had informed his lufthansa training school of an episode of severe depression , airline says .

.

V. CONCLUSION

This article presents a review of abstractive text summarization by surveying transformer related literature. We have implemented four different models namely Transformer, Fastformer, Transformer with Pointer generator networks and coverage, Fastformer with pointer generator networks and coverage. We have demonstrated that each of these models (in this order) outperforms the prior models. For comparative research, we calculated ROUGE ratings for each model's predictions and concluded that the Fastformer model with pointer generator networks and coverage improved the results than other models as Fastformer has linear complexity. Future work should focus on increasing hyper-parameter adjustment and giving each model more time to train. Applying it to different data sets would be a fascinating exercise.

REFERENCES

[1] N. Andhale and L. A. Bewoor, "An overview of text summarization techniques", 2016 International Conferecne on Computing Communication Control and Automation (ICCUBEA), DOI:10.1109/ICCUBEA.2016.7860024.

[2] Madhuri P. Karnik, Dr.D.V.Kodavade, "A Discussion on Various Methods in Automatic Abstractive Text Summarization", International conference on Mobile computing and sustainable Informatics (ICMCSI-2020).

[3] Alexander M. Rush, Sumit Chopra, Jason Weston, "A neural attention model for abstractive sentence summarization". Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 379-389, September 2015.

[4] Sumit Chopra, Michael Auli, Alexander M. Rush, "Abstractive sentence summarization with attentive recurrent neural networks", proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human language Technologies, , pages 93-98, June 2016.

[5] R. Nallapati, B. Zhou, Cicero dos Santos, C. Gulcehre, B. Xiang , "Abstractive text summarization using sequence to sequence RNNs and beyond," in Proceedings of the CoNLL-16, Berlin, Germany, August 2016.

[6] Junyang Lin, Xu Sun, Shuming Ma, Qi Su, "Global Encoding for Abstractive Summarization", In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (pp. 163-169), July 2018.

[7] Shengli Song,Haitao Huang, Tongxiao Ruan, "Abstractive text summarization using LSTM-CNN based deep learning", Multimedia Tools and Applications, vol.-78, issue 1, pp. 857-875, Jan 2018.

[8] Piji Li, Wai Lam, Lidong Bing, Zihao Wang, "Deep recurrent generative decoder for abstractive text summarization", Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 2091–2100, September 2017.

[9] Tal Baumel, Matan Eyal, Michael Elhadad, "Query Focused Abstractive Summarization:Incorporating Query Relevance, Multi-Document Coverage, and Summary Length Constraints into seq2seq Models", arXiv:1801.07704v2 [cs.CL] 25 Jan 2018.

[10] Abu Kaisar Mohammad Masum, Sheikh Abujar , Md Ashraful Islam Talukder , A.K.M.Shahariar Azad Rabby , Syed Akhter Hossain , "Abstractive method of text summarization with sequence to

sequence RNNs", 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need", 31st Conference on Neural Information Processing Systems (NIPS), 2017.

[12] Abigail See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks", Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, vol. 1, July 2017, pp. 1073–1083.

[13] Qian guo, J Huang , N Xiong , P Wang (2019), "MS-Pointer Network: Abstractive Text Summary Based on Multi-Head Self-Attention", volume 7, 2019.

[14] Xin Liu, Liutong Xv , "Abstract summarization based on the combination of transformer and LSTM", 2019 International Conference on Intelligent Computing, Automation and Systems(ICICAS),DOI: 10.1109/ICICAS48597.2019.00199.

[15] Xin Liu, Liutong Xu , "A Combined Model for Extractive and Abstractive Summarization Based on Transformer Model", DOI reference number: 10.18293/SEKE2020-069.

[16] Ekaterina Z, Tsegaye Misikir Tashu and Tomáš Horváth, "Abstractive Text Summarization using Transfer Learning", CEUR- WS. Org /vo1-2718, 2020.

[17] Jon Deaton, Austin Jacobs, Kathleen Kenealy, Abigail See, "Transformers and Pointer-Generator Networks for Abstractive Summarization", 2019.

[18] Wen Kai, Zhou Lingyu , "Research on Text Summary Generation Based on Bidirectional Encoder Representation from Transformers", 2nd International Conference on Information Technology and Computer Application (ITCA), 2020, DOI:10.1109/ITCA52113.2020.00074.

[19] Jianwen J, Haiyang Z, Chenxu D, Qingjuan Z , Hao F , Zhanlin J, "Enhancements of Attention-Based Bidirectional LSTM for Hybrid Automatic Text Summarization", 2021, DOI: 10.1109/ACCESS.2021.3110143.

[20] Jiaming Sun, Yunli Wang, Zhoujun L, "An Improved Template Representation-basedTransformer for Abstractive Text Summarization", 2020 International Joint Conference on Neural Networks (IJCNN), DOI:10.1109/IJCNN48605.2020.9207609.

[21] Soheil Esmaeilzadeh , Gao Xian Peh , Angela Xu , "Neural Abstractive Text Summarization and Fake News Detection" arXiv:1904.00788v2 [cs.CL] 12 Dec 2019.

[22] Urvashi Khandelwal , Kevin Clark, Dan Jurafsky, Łukasz Kaiser , "Sample Efficient Text Summarization Using a Single Pre-Trained Transformer", arXiv:1905.08836v1[cs.CL], 2019.

[23] Vrinda Vasavada, Alexandre Bucquet , "Just News It: Abstractive Text Summarization with a Pointer-Generator Transformer", CS224N: Natural Language Processing with Deep Learning, Stanford University, 2019.

[24] DANG Trung Anh, Nguyen Thi Thu Trang , "Abstractive Text Summarization Using Pointer-Generator Networks with Pre-trained Word Embedding", SoICT'19: Proceedings of the 10th International Symposium on Information and Communication Technology, Pages 473-478, December 2019.

[25] Hangbo Bao, Li Dong, Wenhui Wang, Nan Yang, Furu Wei , "s2s-ft: Fine-Tuning Pretrained Transformer Encoders for Sequence-to-Sequence Learning", arXiv:2110.13640v1 [cs.CL],26 Oct 2021.

[26] Sandeep Subramanian, Raymond Li, Jonathan Pilault, Christopher Pal (2020), "On Extractive and Abstractive Neural Document Summarization with Transformer Language Models", 2020, Conference on Empirical Methods in Natural Language Processing, Nov. 16-20.

[27] Tian Cai, Mengjun Shen, Huailiang Peng, Lei Jiang, Qiong Dai, "Improving Transformer with Sequential Context Representations for Abstractive Text Summarization", Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019. Pages 512-524.

[28] Nima Sanjabi (2018). Abstractive text summarization with attention-based mechanism, Master's thesis, Universitat Politècnica de Catalunya, July 2018.

[29] Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly, "Pointer networks", NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2, Pages 2692-2700, December 2015.

[30] Siddhant Porwal, Laxmi Bewoor, Vivek Deshpande, "Transformer Based Implementation for Automatic Book Summarization", International Journal of Intelligent Systems and applications in engineering (IJISAE), 10(3s), 123-128, 2022.

[31] Prottyee Howlader, Prapti Paul, Meghana Madavi, Dr. Laxmi Bewoor, Dr.V.S.Deshpande, International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 10 Issue: 1s, 2022.

[32] Chuhan Wu, Fangzhao Wu, Tao Qi, Yongfeng Huang, Xing Xie , "Fastformer: Additive Attention Can Be All You Need", arXiv:2108.09084v6 [cs.CL] 5 Sep 2021.