iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61650

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Next-Gen Linguistic Intelligence Image Summarizer for Documents Using Attention Model

Dr. Karthik Elangovan[1], R.S Tabares Ahamed[2], J Somesh[3], Afjal SK[4]
Computer Science and Engineering, SRM Institute of Science and Technology Ramapuram, Chennai -600089, India

Abstract: The process of producing a succinct synopsis of a text involves two primary techniques: both abstractive and extractive While abstractive summarising rephrases text in a way that is more human-like, extractive summarization chooses important statements straight. handbook Because it takes time and money to summarise because it requires reading through big texts, an automated text model with lots of features It is suggested to use both extractive and abstractive methods for summarising. Important sentences that are enhanced by linguistic features are identified by a feature-rich extractor. These sentences are then further enhanced by an abstracter that adds components such named entity tags and part-of-speech tags. Moreover, a loss function is presented to match attentions at the word and sentence levels. At last the summarized text is converted into image by using parsing and other relatable techniques.

Keywords: Abstractive summarization, encoder-decoder, extractive summarization, feature rich model, linguistic features, summarization evaluation, Visual Representation, Application Programming Interface.

I. INTRODUCTION

In the digital era, the exponential growth of textual data presents a significant challenge: how to effectively manage, analyse, and distill this vast amount of information. While traditional text summarization methods have been valuable, they often struggle to capture the intricate essence of complex documents and present information in visually engaging ways.

Recognizing this challenge, our project embarks on an innovative endeavor: the creation of a Next-Generation Linguistic Intelligence Image Summarizer for Documents using an Attention Model. At its core, our project seeks to redefine the landscape of textual information summarization and presentation by leveraging advanced attention mechanisms and sophisticated linguistic intelligence algorithms. Unlike conventional approaches that simply produce textual summaries, our novel method takes a significant leap forward by converting summarized text into visually intuitive representations. By doing so, our aim is to offer users a more immersive and efficient way to comprehend intricate documents. A central aspect of our project involves integrating advanced attention mechanisms, which enable our system to dynamically focus on the most relevant aspects of the input text. This ensures that the resulting visual representations not only capture the essence of the document but also do so in a coherent and meaningful manner. Furthermore, by incorporating linguistic intelligence algorithms, our system goes beyond surface-level summarization to extract semantic meaning and context from the text. This enriches the visual summaries, providing users with deeper insights into the content. Through the synergistic combination of attention mechanisms and linguistic intelligence, our project strives to bridge the gap between textual information and visual comprehension. Our ultimate goal is to empower individuals and organizations with a fresh perspective on document summarization—one that not only enhances efficiency but also fosters deeper engagement and understanding. Our project marks a step, in the realm of processing textual information by converting text into visual form. Through rethinking how information is summarized and presented our goal is to provide users with the tools they need to navigate the world of data in today's digital era all presented in an engaging visual format.

II. LITERATURE SURVEY

In recent years, much of the focus in text summarization has revolved around extractive techniques, where sentences and phrases are identified in a document and reproduced as a text summary again. Numerous research have evaluated the quality of automatic text summarization systems using datasets, attention models, and assessment approaches. [34] used neural networks to convert phrases into vectors, whereas Nallapati et al. [21] and Cheng and Lapata [3] used RNNs to generate document representations. Narayan et al. [23] used a sentence classifier to choose sentences based on titles and picture captions. Yasunaga et al. [33] used graph convolutional networks and RNNs to evaluate sentence significance. Readability problems frequently plague extractive summarization methods, even if some of them get excellent ROUGE ratings.

In contrast to neural models, abstractive document summarization has not, until recently, gotten enough attention.. Jing [15] was among the first to create summaries by eliminating unimportant parts of sentences. The task of creating summaries using methods became standardized during the DUC 2003 and 2004 competitions. TOPIARY [35] found success, in the challenge by using compression techniques and detection algorithms based on language principles. They also added keywords from a document to the result. Cheung and Penn [4] fused sentences using dependency trees, while Rush et al. [25] We present a current neural network for abstractive text summarization that uses convolutional models for text encoding and an attentional feed forward neural network for summary synthesis. Vinyals et al. [28] developed a pointer network using the soft attention distribution approach of Bahdanau et al. [2], which has since been applied to various tasks including language modelling, neural machine translation [11], and summarization [16], [21]. An extension of this work by Rush et al. [25] replaced the decoder with an RNN to improve performance. Hu et al. [13] demonstrated promising results in Chinese dataset summarization using RNNs, while Cheng and Lapata [3] employed an RNN-based encoder-decoder for extractive text summarization. Nallapati et al. [21] utilized a sequence-to-sequence model for summarization evaluation on the CNN/DailyMail dataset. Ranzato et al. [24] replaced the traditional training matrix with evaluation matrices such as ROUGE and BLEU, while See et al. [26] and Jin et al. [16] adopted pointer networks to handle out-of-vocabulary words in their models. Yadav et al. [7] used reinforcement learning with an attention layer, whereas Li et al. [17] used generative adversarial networks to get high ratings in human assessments. Bahdanau et al.[1] proposed an attention mechanism, and Yang et al. [32] introduced a hierarchical attention mechanism for document classification. In a study, by Nallapati and colleagues [21] they combined attention mechanisms at both the word and sentence levels incorporating sentence attention. Furthermore there have been advancements in feature engineering for selecting features in models with a focus on meta learning [36] [37] [38] particularly emphasizing learning and adaptation even in cases of limited data availability. In our research project we introduce a model for summarizing text that involves extracting key information before generating summaries. The words are encoded with features to retain details and are then inputted into both the extractor and summarizer components. This model combines sentence level summarization from an extraction based approach and word level summarization from a generation based method. While past studies have explored attention mechanisms, the incorporation of attention to attributes within a unified framework is innovative. By utilizing word and sentence level attention within a model that encodes words, with features we aim to enhance the comprehensiveness of text summarization.

III.APPROCH

Our solution combines the power of a cutting-edge extractor [22] with a feature-rich abstracter [21]. Throughout this work, we will use the following notation. The extractor and abstractor analyse a word sequence expressed as w = {w1, w2,...,wi,...,wn}, where i represents the word index. These words come together to form sentences denoted as s = {s1, s2,...,sj,...,sn} where j signifies the sentence index. Each word is linked to the sentence based on a mapping function j. Extractive and abstractive summarization entail finding phrases and words in a text, using attention mechanisms to construct a succinct summary. The extractor uses sentence-level attention α = {α1, αm...} to assess the possibility of extracting each sentence into the summary. The abstracter dynamically applies word-level attention (β = βt 1, βt 2,..., βt n). Generate each word at each step t for the summary.

A. Pre-Processing

Texts pre processing aims to reduce ambiguities caused by various verb forms or singular/plural variations. Additionally, stop words like "a," "the," "of," and "is" carry minimal information for summarization purposes. Below are the pre-processing operations applied to documents:

1) *Document Segmentation*: Text is segmented into paragraphs to identify the placement of each sentence within its respective paragraph.
   a) *Stemming:* Stemming is employed to reduce words to their root or base form, such as using singular forms and removing verb endings like "-ing." We utilize the StanfordNLP stemmer for this task.
   b) *Paragraph Segmentation:* Paragraphs are divided into individual sentences using sentence tags.
   c) *Words standardization:* In every sentence words are made consistent by being reduced to their base forms through lemmatization and stemming using Porter’s algorithm [29].
   d) *Stop Word Filtering:* Stop words, such as "a," "is," "in," "the," and "of," are filtered out from the document after other pre-processing steps. Stop word filtering is a standard practice in text mining applications.
   e) *Text Analysis:* Meaningful information is extracted from text data through various techniques, including Part-of-Speech (POS) Tagging, Named Entity Recognition (NER), Dependency Parsing, and Sentiment Analysis.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

f) Image Preprocessing: Image data is prepared for conversion by resizing images to a consistent size, normalizing pixel values, and applying data augmentation techniques like rotation, flipping, and cropping to enhance diversity and robustness.

B. Feature Rich Extractor Model For Extractive Summarization

After simplifying and clarifying a document it is organized into a matrix based on sentence features. Each sentence is analyzed for its features, which are then combined to create a matrix. Through testing features we have determined the set of sentence characteristics to use for the extraction model.

1) Sentence Position

Where a sentence is positioned in a document plays a role, in how the document can be compressed. Past approaches have relied on the sentences placement as it usually contains information. Typically the opening sentences are most vital for summarizing with their impact diminishing as you move away from the beginning of the document.

2) Counting Numbers

2. Numbers often provide information. We determine the ratio of numbers to words, in every sentence using Equation (1).

$sen_{num} = \#numerals / total\#words$

a) Model Architecture

The extractor model builds upon the work of Nallapati et al. [22] but with a different architecture focused on generating highly-ranked sentences based on informativeness derived from recall scores. Hierarchical bidirectional LSTM (Long Short-Term Memory) units are employed, followed by a classification layer to compute sentence-level attention αn (refer to Figure 1). The extractor loss Lossext is computed using a sigmoid cross-entropy function (Equation 2), where K represents the total number of sentences and gk denotes the ground truth for the kth sentence. The model emphasizes recall scores to select highly informative sentences, aiming to include relevant information necessary for abstractive summarization. Ground truth labels g are determined for each sentence to minimize the loss (Equation 3).

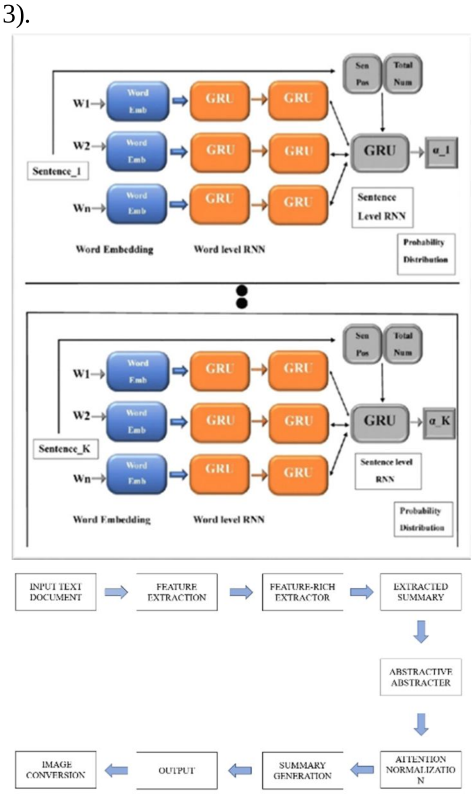


Figure 1: Shows the feature-rich extractor model for n sentences. Long short-term memory (LSTM) with hierarchical bidirectional architecture is used in the model architecture to extract sentence-level data. A classification layer for calculating sentence-level attentiveness αn comes after it.

GRUs (gated recurrent units) [31] extract sentences and distribute attention at the classification layer (see Figure 1). The extractor loss is estimated using a sigmoid cross-entropy function.Equation 2 provides the loss extension: gk, representing the ground truth for the kth sentence, might be 0 or 1, and K is the total number of sentences. A value of gk = 1 indicates that the kth phrase, in a document contributes to summarization. The extractor selects sentences that contain relevant details necessary for generating abstractive summaries.

$$Loss_{ext} = -\frac{1}{K}\sum_{k=1}^{K}((g_k log\alpha_k) + (1-g_k)log(1-\alpha_k)) \quad (2)$$

Sentence extraction is carried out by GRUs (gated recurrent units) [31], which allocate sentence attention αn at the classification layer (refer to Figure 1). Utilizing a sigmoid cross entropy function, the extractor loss is calculated. Equation 2 gives the loss extension: where gk, which represents the ground truth for the kth sentence, is either 0 or 1, and K is the total number of sentences. The kth phrase of a document is narrated by gk = 1, which is noted to facilitate abstractive summarization. Highly informative sentences are chosen by the extractor. This means that some phrases should provide pertinent details needed to create abstract summaries.

b) Weighted Attention

The attention mechanism impacts NLP. A basic solution is to combine word-level attention (βTn) with sentence-level attention (αm) by re-normalization and scalar multiplication. Multiplication is performed where both word-level and sentence-level attention are high. Our focus is on updating word-level attention to enhance abstractive summarization, ensuring consistency between word-level and sentence-level attention during training. We introduce the inconsistent loss Lossics (Equation 3) to reinforce word-level attention when sentence-level attention is high. Various loss functions are applied to the extractor. Abstracter to avoid a solution and support both stages in our suggested two stage integrated model.

$$Loss_{ics} = -\frac{1}{N}\sum_{T=1}^{N} log\left(\frac{1}{|W|}\sum_{n\in W}^{max}\beta_n^T\alpha_{m(n)}\right)$$

Where, W stands for the words that are most frequently used, and N stands for the total amount of words in a given phrase. When sentence-level attention is strong, this strengthens word-level focus. To prevent a deteriorated solution when both word-level and sentence-level attention are high, we employ separate loss functions for the extractor and abstracter. Our proposed two-stage unified strategy utilises an inconsistency loss function that benefits both the extractor and abstractor.

C. Feature Rich Abstracter Model For Abstractive Summarization

Significant language characteristics are computed following pre-processing. The following word characteristics have been shown to be most important for summary.

1) Ne (Named Entity) Tags

Proper nouns in a text string (sentence, paragraph, etc.) are identified by NE tagging techniques. Sentences that relate to name identities—such as a company's name and a personal name—are crucial for providing an accurate description. The place, person, date, organisation, money, percent, and time are the seven classes that make up NE tags. We employ the Stanford NEtagger in this work.

2) Pos Labels

Words in a text are annotated and categorised according to their POS (Parts-of-Speech) categories, which include verbs, adverbs, and nouns. POS tagging may be done using a variety of methods, including statistical techniques like a hidden den Markov model. We employ the Stanford POS tagger in this work.3.

3) No.of Proper Nouns

In corpus linguistics, a word is marked in a text that depends on a certain part of speech. It is not only based on its definition but also on the context, for example, the relationship of a word with its adjacent words in the paragraph. POS tagging identifies words as a noun, pronouns, verbs, adverbs, adjectives, etc. POS tagging is performed based on hidden parts of speech and discrete terms. There are rule-based and statistical approaches. Brill tagger is a rule-based algorithm and is a widely used English POS tagger. This feature is used to count the words in the sentences, which have a considerable number of proper nouns. To compute a few proper nouns, the Stanford POS tagger is used.

4) Term Weights

Another crucial component for handling text summarization is term weights. The term frequency TF indicates a word's significance inside the relevant material. The number of times a word appears in the document is counted and then divided by its length for normalisation, as specified by Equation (4).

$$TF(n) = \frac{\#times\ term\ n\ appear\ in\ the\ document}{total\ \#\ of\ \ terms\ in\ the\ document}$$

The prevalence of inverted documents IDF calculates a term's relative relevance within the text. Equation (5) illustrates that phrases that appear infrequently are typically dragged down and are hence less relevant (scaled up) than ones that occur frequently.

$$IDF\ (n) = log_e \frac{total\ \#\ documents}{\#documents\ having\ term\ n}$$

Lastly, the weight of the phrase is computed using TF × IDF.

5) Use of Model Architecture

Finding the core ideas and central characters that a tale revolves around is a difficult assignment for text summarising. We achieve this by using linguistic characteristics from an input document, such as word weights, POS tags, and NE tags. To preserve the lexical features of words, we provide appended look-up-based vector embedding.

In order to incorporate linguistic information, POS tags substitute grammatical tags for textual descriptions. Discretization converts continuous characteristics like TF and IDF into explicit values. The bin value to which they belong is represented by a single hot encoding of their values. As a result, each word in the look-up dictionary has a word embedding and four linguistic features—POS tags, NE tags, TF, and IDF. Tags and hot encoding are added to the word embedding to create a single, lengthy vector. Only word embedding is provided as input to the decoder. Figure 2 depicts the design of a feature-rich abstract model that generates a detailed summary using accurate linguistic input. The main issue with employing sequence-to-sequence models to generate a summary that consists of several phrases is repetition. In order to tackle this issue, we have embraced the coverage model, wherein the integrated impact of attention distributions calculated from earlier time steps is represented by a coverage vector v t in Equation (6).

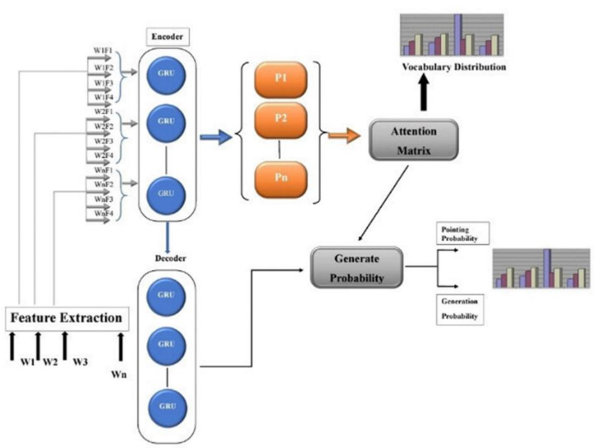


Figure 2. An abstracter model with rich features for n words The four linguistic features—term weight, POS tag, NE tag, and the quantity of proper nouns—that are calculated for every word are represented by the letters F1, F2, F3, and F4. Word embedding is used to concatenate them before feeding them to the abstractor network.

vT shows how the words have been divided from the original material and demonstrates how the attention mechanism has been used to choose these terms. Since no source documents were hidden at the initial time step, v0 here stands for the vector with zero degrees. Equation (7) presents the attention mesh and ism that we added the coverage vector to.

$$E_j^T = V^t tanh(w_H H_j + w_k K_T + W_{cv} C_j^T)$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

where Wcv is a trainable parameter with the vector v's length. It suggests that the previous decisions have provided the attention mechanism for the current decision. By preventing repeated phrases from coming from the same place, this supports the attention process. The act of visiting the source document's location has been made illegal by using coverage loss. The machine translation loss is not the same as the bounded coverage loss. In machine translation, there is a translation ratio; the translation ratio determines the final convergence vector's penalty. The loss function that is being used can be adjusted, as summarization does not need continuous coverage. Penalising the overlapping attention distribution and coverage to prevent reciprocal attention is the aim of the coverage loss function.

IV. EXPERIMENTAL RESULTS

A. Dataset

We evaluated the technique using the CNN/Daily Mail dataset, which included online news stories with an average of 780.9999 tokens. These articles have multi-sentence summaries, with an average of 57 tokens (3.7499 sentences). The dataset is organized into 13,777 validation pairs, 27,777 training pairs, and 11,777 test pairings.

B. The Experiments Procedure

128-dimensional word embeddings were used to train the abstracter and the extractor. We followed See et al. [26] and Nallapati et al. [22] by using 200 and 256 hidden states for the extractor and abstracter, respectively. Nallapati et al. [21] used a vocabulary size of 50,000 words. enabling handling of out-of-vocabulary words. The coverage structure and the pointer generator gave the network a minimum of 512 and 1153 trainable parameters, respectively. We used embeddings learnt during network training, rather than pre-trained embeddings [21]. Both abstractor and extractor have a learning rate of 0.16. Using the Adagrad optimizer [8], the accumulator was set to 0.1. Early stopping was used to cease network training when overfitting occurred on the validation set. The length of the source text was restricted to word tokens for training and testing, with a maximum length of 100 tokens for reference summaries. During testing, 120 tokens were decoded for comparison. To accelerate network training, we shortened the initial encoding and decoding steps to 101 and 51 tokens, respectively. The process of progressively lengthening articles until convergence was sped up by truncating them at the start of training. Furthermore, the model underwent 48,000 iterations of training with a batch size of 4. During the training of the vector representation of words, the linguistic aspects of words were concatenated with word embedding, which did not impact the training time of the main network. The end-to-end model lowered training time by minimising encoding and decoding phases. While the abstracter tried to minimise loss functions with certain values, the extractor was taught to shorten the text. The extractor and abstracter worked together to create a two-stage network in which the pre-trained extractor used sentence-level attention.

C. Findings And Conversation

The network took six days and eighteen hours to train the extractor and abstracter on an 11GB GPU with a batch size of four. After around 118 thousand iterations, the coverage mechanism was introduced, and accuracy rose rapidly throughout the early training phases. After adding coverage with a weighted coverage loss value, training continued for about 2,000 additional iterations (4 hours), leading to improved performance during testing. The network exhibited improved performance, measured by ROUGE scores [19], compared to existing approaches. The proposed two-stage approach outperformed previous work significantly, with the addition of a coverage function reducing redundancy in generated summaries.

TABLE 1 Compares the Nallapati et al. [21] model, the pointer generator [26], and the two-stage network. F-measures for unigrams (ROUGE-1), bigrams (ROUGE-2) and longest common subsequences (ROUGE-L) are shown. Scores were calculated using the rougelib library.

Model	ROUGE-1	ROUGE-2	ROUGE-L
Seq-seq + attention baseline (150k vocab)	30.49	11.17	28.08
Seq-seq + attention baseline (50k vocab)	31.33	11.81	28.83
Pointer Generator	36.44	15.66	33.42
Two-stage network (proposed)	**37.76**	**17.81**	**33.83**

The table shows how much better the suggested two-stage method with the extractor and abstracter performed than the previous effort. Subsequent tests revealed that this enormous vocabulary of 150,000 words did not seem to be helpful for increasing efficiency. Some important information was missed by the performance efficient model. Extensive descriptions were deemed unnecessary for the majority of produced summaries, and uncommon terms were commonly substituted with more common ones. Even though OOV terms were handled using a large vocabulary technique (LVT), we still discovered word duplication in the resulting summaries. However, it was found that adding the coverage function decreased the amount of repeated words. Reducing the duplication issue came at a 1.7% cost. The feature-rich extractor summary includes the following additional information: First, read without reading the original content. Signature of Saili information Head coach Anthony addresses Saili's teammates. Second Reading: Article duration: around 2.5 minutes (full understanding). Extractive Summary of Reference: 30 seconds Feature-rich extracted summary takes around 50-1 minute.Reading for the first time Reference abstractive summary: Understanding the context of an article: On average. Characteristic Abstract and detailed synopsis.

article (truncated): munster have signed new zealand international francis saili on a two year deal. utility back saili, who made his all blacks debut against argentina in 2013, will move to the province later this year after the completion of his 2015 contractual commitments. the 24 year old currently plays for auckland based super rugby side the blues and was part of the new zealand under 20 side that won the junior world championship in italy in 2011. saili's signature is something of a coup for munster and head coach anthony foley believes he will be a great addition to their backline. francis saili has signed a two year deal to join munster and will link up with them later this year. 'we are really pleased that francis has committed his future to the province, 'foley told munster's official website. 'he is a talented centre with an impressive skill set and he possesses the physical attributes to excel in the northern hemisphere. 'i believe he will be a great addition to our backline and we look forward to welcoming him to munster. 'saili has been capped twice by new zealand and was part of the under 20 side that won the junior championship in 2011.
reference extractive summary: munster have signed new zealand international francis saili on a two year deal. utility back saili, who made his all blacks debut against argentina in 2013, will move to the province later this year after the completion of his 2015 contractual commitments. the 24 year old currently plays for auckland based super rugby side the blues and was part of the new zealand under 20 side that won the junior world championship in italy in 2011. francis saili has signed a two year deal to join munster and will link up with them later this year. 'this experience will stand to me as a player and i believe i can continue to improve and grow within the munster set up.
reference abstractive summary: utility back francis saili will join up with munster later this year. the new zealand international has signed a two year contract. saili made his debut for the all blacks against argentina in 2013.
feature rich extractor summary: munster have signed new zealand international francis saili on a two year deal. utility back saili, who made his all blacks debut against argentina in 2013, will move to the province later this year after the completion of his 2015 contractual commitments. the 24 year old currently plays for auckland based super rugby side the blues and was part of the new zealand under 20 side that won the junior world championship in italy in 2011. saili's signature is something of a coup for munster and head coach anthony foley believes he will be a great addition to their backline. francis saili has signed a two year deal to join munster and will link up with them later this year. saili has been capped twice by new zealand and was part of the under 20 side that won the junior championship in 2011. saili, who joins all black team mates dan carter, ma'a nonu, conrad smith and charles piutau in agreeing to play his trade in the northern hemisphere, is looking forward to a fresh challenge. he said: 'i believe this is a fantastic opportunity for me and i am fortunate to move to a club held in such high regard, with values and traditions i can relate to from my time here in the blues.
feature rich abstracter summary: munster have signed new zealand international francis saili on a two year deal. utility back saili made his all blacks debut against argentina in 2013. the 24 year old currently plays for auckland based super rugby side the blues.

Figure 3. Compares extractive and abstractive summaries of a news story prepared using the suggested methodology.

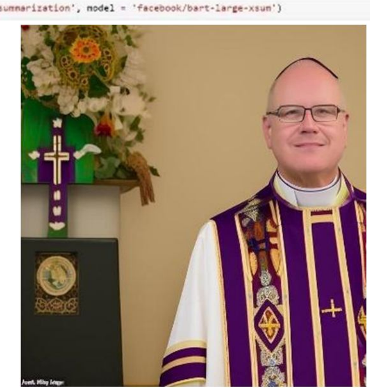
```
    num_rows: 100
})


Dataset({
    features: ['id', 'article', 'highlights'],
    num_rows: 100
})

In [59]: train_ds[0]

Out[59]: {'id': '0001d1afc246a7964130f43ae940af6bc6c57f01',
 'article': "By . Associated Press . PUBLISHED: . 14:11 EST, 25 October 2013 . | . UPDATED: . 15:36 EST, 25 October 2013 . The bishop of the Fargo Catholic Diocese in North Dakota has exposed potentially hundreds of church members in Fargo, Grand Forks and Jamestown to the hepatitis A virus in late September and early October. The state Health Department has issued an advisory of exposure for anyone who attended five churches and took communion. Bishop John Folda (pictured) of the Fargo Catholic Diocese in North Dakota has exposed potentially hundreds of church members in Fargo, Grand Forks and Jamestown to the hepatitis A . State Immunization Program Manager Molly Howell says the risk is low, but officials feel it's important to alert people to the possible exposure. The diocese announced on Monday that Bishop John Folda is taking time off after being diagnosed with hepatitis A. The diocese says he contracted the infection through contaminated food while attending a conference for newly ordained bishops in Italy last month. Symptoms of hepatitis A include fever, tiredness, loss of appetite, nausea and abdominal discomfort. Fargo Catholic Diocese in North Dakota (pictured) is where the bishop is located .",
 'highlights': 'Bishop John Folda, of North Dakota, is taking time off after being diagnosed .\nHe contracted the infection through contaminated food in Italy .\nChurch members in Fargo, Grand Forks and Jamestown could have been exposed .'}

In [ ]: summarizer = pipeline('summarization', model = 'facebook/bart-large-xsum')
```



Text Output

Table 2 compares NALLAPATI ET AL. [21], POINTER GENERATOR [26], THE TWO-STAGE NETWORK (PROPOSED), and THE REFERENCE SUMMARY.

Model	Comprehensiveness	Conciseness	Informativeness
Nallapati et. al	3.49	2.64	3.17
Pointer Generator	3.53	2.84	3.34
Two-stage network (proposed)	**3.89**	**2.97**	**3.62**

Excellent understanding of the article's background. Finally, the feature-rich model is crucial for retaining important proper nouns, numerals, and phrases in the summary. The created summary considerably reduced the time required to read the complete information.

D. Assessment of Humans

Amazon Mechanical Turk was used for the human evaluation, and 50 test samples were chosen at random for analysis. Each sample consisted of an original article, baseline, two-stage network output, and reference summaries. On a scale of 1 to 5, evaluators graded summaries for comprehensiveness, conciseness, and informativeness. In comparison to more contemporary summarising algorithms, the two-stage model performed well, especially in terms of comprehensiveness, because it could extract information based on linguistic features. This led to higher ratings.

V. CONCLUSION

In this study, we introduced a novel approach that leverages the strengths of both extractive and abstractive summarization models to produce comprehensive summaries. Our neural network effectively extracts the main ideas from the original text by integrating a variety of linguistic variables into the word embeddings, including sentence position, numbers, POS tags, NE tags, term weights, and proper nouns.Through attention mechanisms at both the sentence and word levels, our model effectively identifies and highlights the most crucial information. Our proposed two-stage model seamlessly integrates extractive and abstractive summarization within a single framework. By training and testing on the CNN/Daily Mail dataset, we demonstrated the effectiveness of our approach, achieving a ROUGE score of 37.76%. Additionally, human evaluation confirmed the high comprehensiveness and informativeness of our generated summaries. Furthermore, we explored the conversion of summarized text into image format using parsing and other techniques. This innovative approach enhances accessibility and visualization of the summarized content, providing a more engaging and intuitive way to consume information. Overall, our study contributes to advancing the field of text summarization by offering a robust and efficient method for generating high-quality summaries, while also exploring novel avenues for presenting summarized content through image conversion techniques.

REFERENCES

[1] D. Bahdanau, K. Cho, and Y. Bengio, ‘‘Neural machine translation by jointly learning to align and translate,’’ 2014, arXiv:1409.0473.
[2] D. Bahdanau, J. Chorowski, D. Serdyuk, P. Brakel, and Y. Bengio, ‘‘End-to-end attention-based large vocabulary speech recognition,’’ in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Mar. 2016, pp. 4945–4949.
[3] J. Cheng and M. Lapata, ‘‘Neural summarization by extracting sentences and words,’’inProc.54thAnnu.MeetingAssoc.Comput.Linguistics(Long Papers), vol. 1, 2016, pp. 484–494.
[4] J. C. K. Cheung and G. Penn, ‘‘Unsupervised sentence enhancement for automatic summarization,’’ in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), Oct. 2014, pp. 775–786.
[5] J. Chorowski, D. Bahdanau, D. Serdyuk, K. Cho, and Y. Bengio, ‘‘Attention-based models for speech recognition,’’ in Proc. 28th Int. Conf. Neural Inf. Process. Syst., vol. 1, Jun. 2015, pp. 577–585.
[6] C. A. Colmenares, M. Litvak, A. Mantrach, and F. Silvestri, ‘‘HEADS: Headline generation as sequence prediction using an abstract feature-rich space,’’ in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol., 2015, pp. 133–142.
[7] A. K. Yadav, A. Singh, M. Dhiman, R. Kaundal, A. Verma, and D. Yadav, ‘‘Extractive text summarization using deep learning approach,’’ Int. J. Inf. Technol., vol. 14, no. 5, pp. 2407–2415, 2022.
[8] J. Duchi, H.Elad, and Y.Singer, ‘‘Adaptive subgradient methods for online learning and stochastic optimization,’’ J. Mach. Learn. Res., vol. 12, no. 7, pp. 1–39, Jul. 2011.
[9] G. Erkan and D. R. Radev, ‘‘LexRank: Graph-based lexical centrality as salience in text summarization,’’ J. Artif. Intell. Res., vol. 22, pp. 457–479, Dec. 2004.
[10] K. Filippova and Y. Altun, ‘‘Overcoming the lack of parallel data in sentence compression,’’ in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), Oct. 2013, pp. 1–11.
[11] C. Gulcehre, O. Firat, K. Xu, K. Cho, L. Barrault, H.-C. Lin, F. Bougares, H. Schwenk, and Y. Bengio, ‘‘On using monolingual corpora in neural machine translation,’’ 2015, arXiv:1503.03535.

[12] W.-T. Hsu, C.-K. Lin, M.-Y. Lee, K. Min, J. Tang, and M. Sun, ‘‘A unified model for extractive and abstractive summarization using inconsistency loss,’’ in Proc. 56th Annu. Meeting Assoc. Comput. Linguistics (Long Papers), vol. 1, 2018, pp. 132–141.

[13] B. Hu, Q. Chen, and F. Zhu, ‘‘LCSTS: A large scale Chinese short text summarization dataset,’’ in Proc. Conf. Empirical Methods Natural Lang. Process., 2015, pp. 1967–1972.

[14] S. Jean, K. Cho, R. Memisevic, and Y. Bengio, ‘‘On using very large target vocabulary for neural machine translation,’’ in Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process., vol. 1, 2015, pp. 1–10.

[15] H. Jing, ‘‘Sentence reduction for automatic text summarization,’’ in Proc. 6th Conf. Appl. natural Lang. Process., 2000, pp. 310–315.

[16] J. Jin, P. Ji, and R. Gu, ‘‘Identifying comparative customer requirements from product online reviews for competitor analysis,’’ Eng. Appl. Artif. Intell., vol. 49, pp. 61–73, Mar. 2016.

[17] J. Li, W. Monroe, T. Shi, S. Jean, A. Ritter, and D. Jurafsky, ‘‘Adversarial learning for neural dialogue generation,’’ 2017, arXiv:1701.06547.

[18] C.-Y. Lin, G. Cao, J. Gao, and J.-Y. Nie, ‘‘An information-theoretic approach to automatic evaluation of summaries,’’ in Proc. Main Conf. Hum. Lang. Technol. Conf. North Amer. Chapter Assoc. Comput. Linguis tics, 2006, pp. 463–470.

[19] C. Y. Lin, ‘‘Looking for a few good metrics: Automatic summarization evaluation-how many samples are enough?’’ in Proc. NTCIR, Jun. 2004, pp. 1–10.

[20] S. Jadooki, D. Mohamad, T. Saba, A. S. Almazyad, and A. Rehman, ‘‘Fused features mining for depth-based hand gesture recognition to clas sify blind human communication,’’ Neural Comput. Appl., vol. 28, no. 11, pp. 3285–3294, 2017.

[21] R. Nallapati, B. Zhou, C. dos Santos, C. Gulcehre, and B. Xiang, ‘‘Abstractive text summarization using sequence-to-sequence RNNs and beyond,’’ in Proc. 20th SIGNLL Conf. Comput. Natural Lang. Learn., 2016, pp. 280–290.

[22] R. Nallapati, F. Zhai, and B. Zhou, ‘‘SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of docu ments,’’ in Proc. 31st AAAI Conf. Artif. Intell., 2017, pp. 3075–3081.

[23] S. Narayan, N. Papasarantopoulos, S. B. Cohen, and M. Lapata, ‘‘Neural extractive summarization with side information,’’ 2017, arXiv:1704.04530.

[24] M. A. Ranzato, S. Chopra, M. Auli, and W. Zaremba, ‘‘Sequence level training with recurrent neural networks,’’ in Proc. 4th Int. Conf. Learn. Represent. (ICLR), 2016, pp. 1–16.

[25] A. M. Rush, S. Chopra, and J. Weston, ‘‘A neural attention model for abstractive sentence summarization,’’ in Proc. Conf. Empirical Methods Natural Lang. Process., 2015, pp. 1–11.

[26] A.See,P.J.Liu,andC.D.Manning,‘‘Gettothepoint:Summarizationwith pointer-generator networks,’’ in Proc. 55th Annu. Meeting Assoc. Comput. Linguistics, vol. 1, 2017, pp. 1073–1083.

[27] S. Venugopalan, M. Rohrbach, J. Donahue, R. Mooney, T. Darrell, and K. Saenko, ‘‘Sequence to sequence—Video to text,’’ in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 4534–4542.

[28] O. Vinyals, M. Fortunato, and N. Jaitly, ‘‘Pointer networks,’’ in Proc. 28th Int. Conf. Neural Inf. Process. Syst., vol. 2, 2015, pp. 2692–2700.

[29] P. Willett, ‘‘The Porter stemming algorithm: Then and now,’’ Program, Electron. Library Inf. Syst., vol. 40, no. 3, pp. 219–223, 2006.

[30] K.-F. Wong, M. Wu, and W. Li, ‘‘Extractive summarization using super vised and semi-supervised learning,’’ in Proc. 22nd Int. Conf. Comput. Linguistics COLING, 2008, pp. 985–992.

[31] J.Amin,M.Sharif,M.Raza,T.Saba,R.Sial,andS.A.Shad,‘‘Brain Tumor detection: A long short-term memory (LSTM)-based learning model,’’ Neural Comput. Appl., vol. 32, no. 20, pp. 15965–15973, Oct. 2020.

[32] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, ‘‘Hierarchical attention networks for document classification,’’ in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol., 2016,pp.1480–1489.

[33] M.Yasunaga,R.Zhang,K.Meelu,A.Pareek,K.Srinivasan,andD.Radev, ‘‘Graph-based neural multi-document summarization,’’ in Proc. 21st Conf. Comput. Natural Lang. Learn. (CoNLL), 2017, pp. 452–462.

[34] X. C. Yin, W. Y. Pei, J. Zhang, and H. W. Hao, ‘‘Multi-orientation scene text detection with adaptive clustering,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 9, pp. 1930–1937, Sep. 2015.

[35] D. Zajic, B. Dorr, and R. Schwartz, ‘‘Topiary,’’ in Proc. HLT NAACL Document Understand. Workshop, Boston, MA, USA, 2004, pp. 112–119.

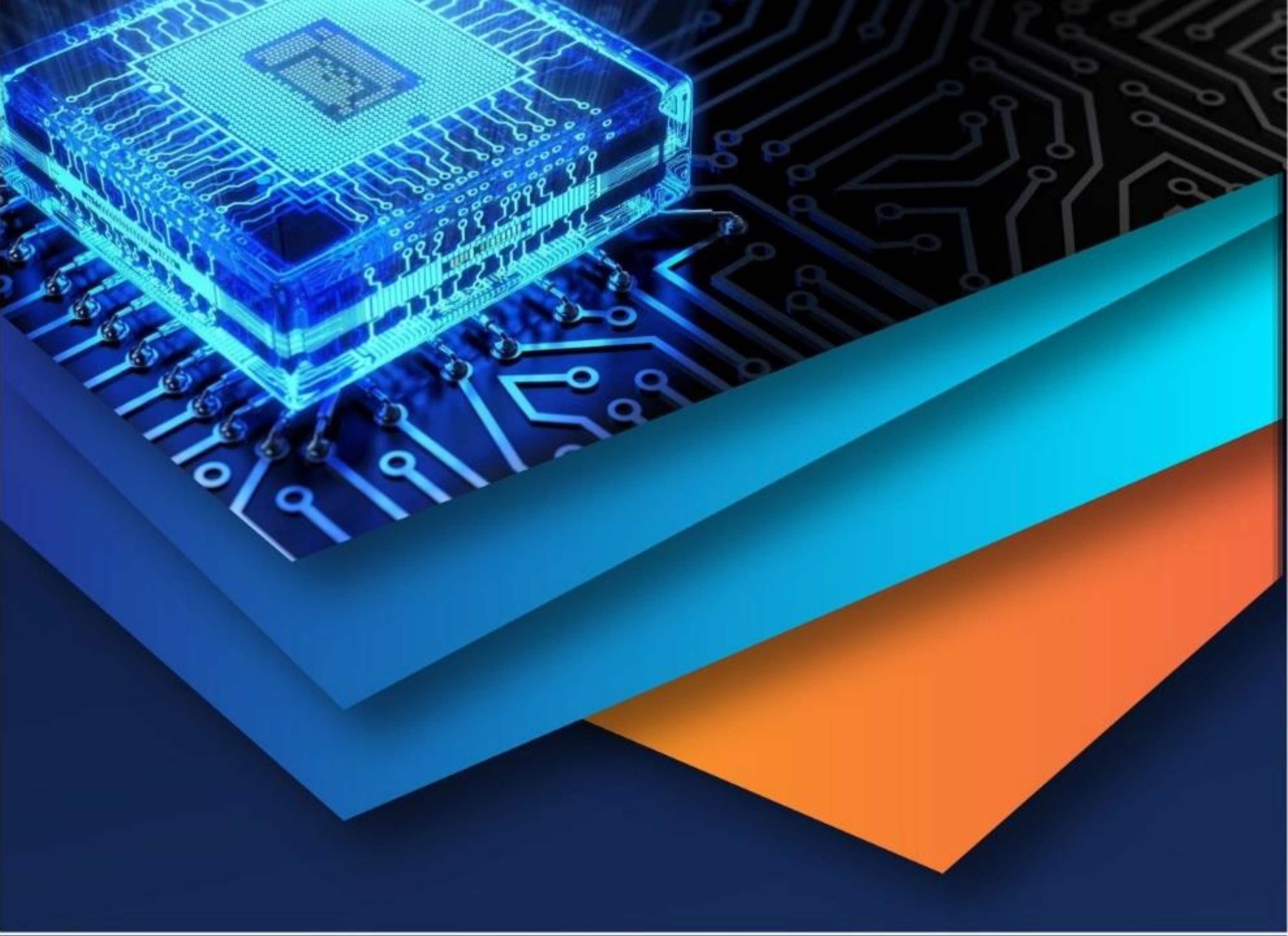
[36] C. M. Bishop, Neural networks for pattern recognition, Oxford university press, 1995.

[37] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization", CVPR, 2007.

[38] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation", CVPR, 2015.

[39] j.Gllavata, R. Ewerth and B. Freisleben, A robust algorithm for text detection in images, pp. 611-616

[40] kumar Anubhav, "An efficient text extraction algorithm in complex images", Contemporary Computing (IC3), 2013



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY