



## Abstractive method-based text summarization using bidirectional long short-term memory and pointer generator mode

Saroj Anand Tripathy<sup>a</sup> • A. Sharmila<sup>b\*</sup>

<sup>a</sup>SENSE, Vellore Institute of Technology Vellore, Tamil Nadu, India

<sup>b</sup>SELECT, Vellore Institute of Technology Vellore, Tamil Nadu, India

Received 02 10 2021; accepted 05 02 2022

Available 02 28 2023

**Abstract:** With the rise of the Internet, we now have a lot of information at our disposal. We are swamped from many sources — news, social media, to name a few, office emails. This paper addresses the problem of reading through such extensive information by summarizing it using text summarizer based on Abstractive Summarization using deep learning models, i.e., using bidirectional Long Short-Term Memory (LSTM) networks and Pointer Generator mode. The LSTM model (which is a modification of the Recurrent Neural Network) is trained and tested on the Amazon Fine Food Review dataset using the Bahadua Attention Model Decoder with the use of Conceptnet Numberbatch embeddings that are very similar and better to GloVe. Pointer Generator mode is trained and tested by the CNN / Daily Mail dataset and the model uses both Decoder and Attention inputs. But due 2 major problems in LSTM model like the inability of the network to copy facts and repetition of words the second method is, i.e., Pointer Generator mode is used. This paper in turn aims to provide an analysis on both the models to provide a better understanding of the working of the models to enable to create a strong text summarizer. The main purpose is to provide reliable summaries of datasets or uploaded files, depending on the choice of the user. Unnecessary sentences will be rejected in order to obtain the most important sentences. The use of different parameters for measuring the performance of the model helped to analyze its efficiency in terms of Cosine Similarity of sentence and Model: 0.29277002,

**Keywords:** Text summarizer, deep learning, bidirectional long short-term memory, pointer generator mode, Bahadua attention model decoder, Conceptnet Numberbatch, GloVe

\*Corresponding author.

E-mail address: asharmila@vit.ac.in (A. Sharmila).

Peer Review under the responsibility of Universidad Nacional Autónoma de México.

## 1. Introduction

Research in the field of text analysis is currently underway. We have two main approaches to summarizing the text.

- Extractive method, which selects specific keywords from the input for the output generation, tends to work with this model, but does not produce structured sentences correctly as it simply selects words from the input and copies them to the output, but does not really understand the phrases, consider it to be a highlight.

- Abstractive method It is building a neural network to really develop the relationship between input and output, not only copy words, it would use this method, it would look like a pen.

The most demanding problem, however, is the abstract summary. This is an area where machine learning has progressed slowly. It's a difficult problem, because creating abstract resumes requires the subject and natural language to be controlled properly, something that can be difficult tasks for a machine.

Encoder — Bidirectional LSTM layer which extracts original text information. The two-headed LSTM reads one word at a time and since it is an LSTM, it updates its hidden state by referring to the word you read earlier and the words you read.

Decoder — LSTM unidirectional layer which produces one word at a time. When the full source text is read, the LSTM decoder begins to operate when the signal is received. It uses encoder information as well as what was previously written to create a probability distribution in the following word.

Attention Mechanism — Encoder and decoder are the building blocks, but the historical decoder's architecture was not very successful by itself. The decoder input without attention is the final hidden status of the encoder, which could be a 256 or a 512-dimensional vector and would become a bottleneck of information if we imagine this small vector could not have the information all in it. The decoder can access the hidden middle states of the encoder via the attention mechanism and use all this information to decide which word to follow.

The above architecture is good enough to begin, as the Pointer generator shows, but two problems are presented in the summaries:

- Summaries reproduce factual details sometimes wrongly (for example Germany beat Argentina 3–2). In rare or non-vocabulary words such as 2–0, this is particularly common.
- Often repeated summaries. (e.g. beats Germany, beats Deutschland etc.)

This is solved by the Pointer Generator model through the creation of a pointer mechanism which enables you to switch between text generation and source copying. Consider the pointer as a probability scale from 0 to 1. If the model is 1, the abstract value is generated, and when it is 0, the word extractive value is generated.

## 2. Literature review

Heading: left justified, in 10-point Arial bold font, sentence-case (capitalization of only the first word, proper nouns and as dictated by other specific English rules; e.g., “The technological history of Japan”). Several studies have been conducted on text synthesizing using various methods, algorithms and models in order to understand how these models are designed to enable both text summarization and text failure, which allows a better path to apply new models. [Madhuri & Kumar \(2019\)](#) introduce a new statistical method in which an extractive text description can be drawn up in a single document. The phrase extraction process is presented which gives a short format to the idea of an input text. Sentences are classified by weight allocation and are classified by weight. High-ranking phrases are removed from the input document in such a way that essential phrases are extracted that produce a high-quality summary of the input document and store the summary as audio.

[Devasena and Hemalatha \(2012\)](#) proposed the research project consisting of automatic categorization of text and a brief analysis approach for the structure of the input text. This ([Devasena & Hemalatha, 2012](#)) work uses a three-stage procedure for reducing rule (TRD), namely Token Creation, Features Identification, Categorization and Summary, to derive the text structure of an input text. This analyzer is tested and gives remarkable results with sample input text. Extensive experiments validate the choice of parameters and the effectiveness of our approach to text classification. This work can also be extended to include indexation for document retrieval, organization and maintenance of large web resources catalogs, automatic metadata extraction, and Word sense ambiguity.

To obtain a briefer summary of the text, [Gunawan et al. \(2019\)](#) conducted research to reduce like phrases from multi-document documents that share similar information. This study uses an online article mix, divided into six categories, to achieve this objective. Combined papers are pre-processed to generate a concise text. After a simple word is received, this study uses the TextRank algorithm to extract important phrases with similarity measures. A summary of the text is provided in this process. However, the summary text also has the same phrases. The next measurement process for reducing similar phrases is the maximum marginal importance (MMR). The outcome of this process is a review of the final text. In the assessment, the average F-score is 0.5103 and 0.4257, respectively ROUGE-1 and ROUGE-2. In this paper, [Reategui et al. \(2012\)](#), discusses a mining tool that can extract text graphs and offers students an abstract for their use. The text summary method relies on the use of graphs as graphic organizers, so that students can reflect on the principal ideas of the text before going on the actual work of writing. The

experiment showed that the tool helped students think about the main ideas of the text and supported the summary writing.

Krishnaveni and Balasundaram (2017) proposed the approach herence and thus enhancing the understandability of the summary text. It summarizes the input document in question with the local value and the local ranking under 'wise summary.' It ranks the phrases wisely and selects the top n phrases from each heading where n depends on the compression ratio. The final heading of the wise summary produced by this approach is a summary of the individual headings. As the 'wise summary' heading has the same proportion of phrases in each heading, the coherent gap in the summary text is reduced. The whole significance and understanding of the abstract text is also improved. The experimental results show clearly that the heading wise synthesizer provides greater accuracy, reminder and f-message through the main synth, the MS-word synthesizer, the free synthesizer and the Auto synthesizer.

Hark et al. (2018) studied texts that do not have a certain structure have been pre-processed and a structured form in expressed form has been translated into the proposed diagram. Different methods could be employed in charts for the extraction of features. Our method conceptually uses the diagrams obtained in the text representation. This study aims to propose a method in which the texts can be summarized by weighing the importance of sentences linearly. In addition, the method presented does not require profound language knowledge and can be adapted to various languages. Mishra et al. (2019) tried to emphasize key techniques for extracting important information from text by modelling the subjects, extracting key sentences and generating a summary. The weighted TF-IDF method is used to extract key phrases for the topic modeling of the LSI and NMF methods, and the text summary is generated with an LSA and Text Rank methods.

Afsharizadeh et al. (2018) created a text summary technique based on the query suggested in paper by extracting the most informative phrases. To this end, a certain number of characteristics are taken from the sentences, each evaluating the importance of the sentences. In this paper we extract 11 of the best features from each sentence. This article showed that using better features results in better abstracts. To evaluate the automatic summaries generated, the ROUGE criterion was used. Ren and Guo (2019) hybridized extractive and abstract resumed systems advantages by offering a combination of Global Gated Unit and copy mechanism (GGUC) text summary model. The experiment results indicate that the model 's performance is superior to the other LCSTS dataset text summary. Chen and Zhuge (2018) proposed a multimodal RNN extractive neural method. Initially, this method encodes multi-modal RNN documents and images, then calculates the overall probability of sentences using a logistic classifier, using text coverage, redundancy and image

coverage as the DailyMail Corps expansion features, by collecting images from the Web. Experiments show that our method is superior to advanced methods of neural resuming. Jiang et al. (2009) proposed 2 approaches in the first approach, the text resumé is used directly to select and categorize characteristics rather than their origin. The second approach is to select and weight each function for each document and to classify free texts using the kNN algorithm. The second approach uses every summary. Experimental results indicated that the two proposed automatic synthetization methods not only can reduce classification times, but also improve text categorization performance.

Zhang and Li (2009) used a summary-based phrase clustering approach. The approach proposed consists of three steps. Firstly, cluster phrases based on the semantic distance between the documents and each cluster will calculate the accumulative similarity of the phrases based on the multi-faceted combination method. The purpose of this paper is to show how the overall outcome depends not only on the characteristics of the sentence, but also on how the sentence is similar. The DUC 2003 data set experimental results showed that our approach could increase performance compared to other summary methods. Jo (2017) proposed that KNN (K Nearest Neighbor) version has been used to calculate the similarity between feature vectors taking into account similarities between attributes or functions and between values. The text summary task shall be viewed in binary classification, where every paragraph or phrase is classed as essence or non-essential and the proposed version improves the results of earlier works by classifying and clustering text. In this research, we identify the similarity between attributes and values, modify the KNN in a similarity version and use the modified version as a summary text approach.

Zenkert et al. (2018) based their paper mining operation is on Multidimensional Knowledge Representation (MKR). The analytic results from different methods of texts mining such as identification, sentiment analysis or subject detection are represented as dimensions of knowledge to support the discovery of knowledge, visualization or computer-aided writing. Summary extractive text is a content-based task that uses the text information that is available to cut down the text length to summarize the text. The knowledge base of the MKR offers an innovative text summary selection tool for this purpose.

Pal and Saha (2014) proposed a technique which performs a summary job with an unsupervised method of learning. With the help of a simplified read algorithm, the significance of the sentence in the input text is assessed. The online semantic dictionary is used as WordNet. The approach proposed gives a 50% overview of the original text and gives a good result, while an original text is 25% overviewed. Rashidghalam et al. (2016) used The BabelNet Knowledge Base and its concept

diagram provide a text summary system. The proposed approach extracts concepts of words that use the Babel Net knowledge base and produces concept charts and evaluates phrases based on the concepts and graphs that result. The final summary therefore uses these rating concepts. Furthermore, an approach to controlling replication is proposed so that selected concepts in each country are punished and result in less redundant summaries. In DUC2004 the performance of the proposed methodology is compared and evaluated, and ROUGE is used as an evaluation metric. The proposed method produces more summaries of the quality and less redundancies compared to other methods. [Alfarra et al. \(2019\)](#) used GFLES to distinguish between its four advantages: 1) by using the graphical model to extract characteristics; 2) by improving substitution for the Vector Space Model (VSM); 3) by taking account of the importance of text subtopics before generating a summary of the clustering phrases; and 4) for both single and several documents. Finding the basis for further development of GFLES to be applied to large data is also the application of the FL on each cluster. Our experimental results show that GFLES can produce extensive summaries of text using the DUC 2004 data set. These findings define the potential of GFLES as an automatic text summary high-precision model. [Hanunggul and Suyanto \(2019\)](#) improved translation of the neural machine (NMT). Two kinds of attention exist: global and local. This article focuses on comparing the effect of local attention in the Long Short-Term Memory (LSTM) model in the case of abstract text resuming (ATS). Developing and evaluating a model with the Amazon Fine Food Review dataset with the GloVe dataset indicates that a global attention model produces better ROUGE-1, which generates more words in the current overview. But ROUGE-2 is higher on the basis of local attention and produces more pairs of words in the current summary, because the local focus mechanism considers the word input subset instead of the input word.

[Iwasaki et al. \(2019\)](#) studied a neural network automatic abstract text summary algorithm developed in Japanese. For experimental purposes we have used a sequence-to-sequence encoder-decoder model. The encoder was supplied with a functional input vector for phrases using BERT. The transforming decoder returned the encoder's summary phrase of the output. This experiment was conducted using the above model with a live-door news corpus. However, there were problems because in the summary sentence the same texts were repeated. [Geng et al. \(2010\)](#) solved the imperfections in the fact that the conventional method for summary text is inappropriate for expressing the meanings of the domain specific documents, the techniques referred to as pos-processing involve: elimination of redundancy, adaptation to rough document summary through clustering paragram(s); generalization of summary phrases. Lastly,

experiments are carried out on clothing field documents and software named CTS is developed according to the above theories. The test results showed that we are efficient and effective in comparison to MS Word 2003. The results of your web pages according to your domain can easily be introduced to mobile devices.

[Ren and Zhang \(2019\)](#) used sequence-to-sequence careful mechanism sequence models have produced good results in abstract text summary. Word vector and spoken vector as model input, and improve the abstract quality of the combined neural convolution (CNN) and bi-directional LSTM. Second, the network of the pointer generator is used to check whether the OOV problem is generated or copied by words. Finally, to monitor the abstract we generated we use the coverage mechanism to avoid problems related to duplication. Our model's ROUGE scores improved significantly and the performance of the LCSTS data packets better than the current state of the art model compared with the classic pointer generator network. [Pattnaik and Nayak \(2019\)](#) in their work use hierarchical clustering effectively by applying cosine-like similarity measures for separating phrases. The model uses an extractive way to summarize the Odia text. It concentrates on reducing redundancies, which it achieves by a cosine matrix of resemblance. Although for European languages such as English, the computerized and complex morphologic structure of the Odia language is primitive, it is a novel work. The results can be modestly successful. [Masum et al. \(2019\)](#) used the amazon fine food review dataset available on Kaggle to make a good summary. The text descriptions of the reviews were used as our input and a simple synopsis of the review descriptions as our output. In order to produce an extensive summary, we have employed a bilateral RNN with LSTM in the layer of the encoding and attention model in the decoding layer. They applied the sequence to the sequence model in order to generate a brief summary of food descriptions. The abstract text synthesizers are challenging such as text processing, speech counting, missing word counting, word embedding, model performance or loss value reduction, and the machine response can be fluent summary. This paper mainly had the objective, in order to provide a better summary of the abstract text, of increasing efficiency and reducing sequence loss by model. They reduced the losing of training by 0.036 successfully in their experiments, and a short summary of the English to English text could be provided through their abstract text summary. [Prakash and Shukla \(2014\)](#) used a single Document in the proposed for the human-aided text summary "SAAR." A term-phrase matrix is generated in the document. Reinforcement learning is the weight of the matrix entries. The summary created is displayed to the user and is the last summary, if approved by the user, otherwise user feedback as keywords will produce a new summary. The results of the DUC2006 documents show that

the performance of the approach proposed in terms of precision, recall and F-score is highly favorable compared to other approaches.

Sherry and Bhatia (2015) used single text summary algorithm proposal for a single document. Text Summary extracts important information from large documents by neglecting the use of different algorithms and providing a very useful compressed overview. Since UNL is an organization which is linguistically independent, UNL summarizes and eliminates unnecessary relationships. UNL is used for text summary because in his native language the user can receive a summary. Lukyamuzi et al. (2019) developed A summary that relies on an improved classification domain-based corpus. Three contributions are made: (1) a synthesis is extended to cover the search for a food security text, (2) a proper corpus is used to support an up-to - date summary, and (3) a less reliable extractive summary can be made in certain conditions. The human task of identifying whether or not there is a food safety article is tested on these study aspirations. Results have shown that 98% under the curve (AUC) after the summary was used a human could still properly label the article, compared with a random rating of 74%. Conversations about food insecurity can easily be identified and monitored in the articles by summarizing them. Trends in such discussions may inform food aid agencies in accordance with prevailing circumstances of suitable measures.

### 3. Methodology

#### 3.1. Model using bi-directional LSTM

We represent in this section our abstract text summarizer methodology for English to English text documents. There have been previously numerous successful works to summarize text in English. However, we have tried to create a better summary text using our own approach and created a better abstract text synopsis for English texts. We have used TensorFlow CPU version to build our model and training.

Fig. 1 displays the working process of my methodology with the use of the block diagram.

#### A. Dataset

The dataset used to develop the model is “Amazon Fine Food Reviews”. Fig. 2 shows the head of the file used. The Dataset has total size of 568,454 with 10 features. It is used in the ratio of (Training Dataset: Testing Dataset: Filtered Dataset) 80:12:8.

#### B. Preprocessing

For preprocessing of the dataset, I had to prepare the dataset because in the model the dataset will be a text with a summary which is converted to vectors using Word2Vec so it can be easily processed by the algorithm. The steps are as follows: 1) Converted all texts to lower case. 2) Several contractions

added (such as ain't, didn't, I'm, etc.). 3) Cleaning text to remove unimportant features and elements from the text by use of the regular expression. 4) Removal of stopwords (words with no or little meaning). 5) Lemmatization of the words to remove synonyms. After this the text is now ready for next processes.

#### C. Further filtration

For shortening training time due to model limits, the dataset is determined by choosing data that does not exceed 300 characters and not less than 25 characters per text. The data set analyzed the data set used, which is 38.15 tokens in the text in the average number of tokens in the dataset and 2.2 tokens in the total dataset summary. After the distribution, the data are processed in a couple of steps consisting of clean text and tokenization, and filtering.

All symbols or characters not necessary on the data are omitted and all letters are used in lowercases during the cleaning text and tokenization process. Then in every text and summary, all the data are converted to tokens. The GloVe dataset refers to word integration in the filtering step. There are several words not registered in GloVe from the word embedding process. In this instance, the model does not pay attention to the word limit, so the whole word is not entered into the word embedding index (GloVe) as <UNK>.

Batch distribution is the last step. The filtering process will be determined by the batch size of the dataset. One text with the largest number of tokens, the number, is selected from a lot. The benchmark. At the end of each summary token, the <EOS> sentence is added. The <PAD> token will include all the text containing the batch until the text is as numerous as the reference text. The same is also true for Every text's summaries. This was to create a fixed length for each array in one batch.

We used a pre-trained word in a vector file to enhance our model. Pre-trained text for vector files like GloVe, ConceptNet Numberbatch, wiki-news-300d-1M.vec, crawl-300d-2M.vec and others are available. We used ConceptNet Numberbatch, based on our work here. We have fixed the size of vocabulary and count the words in our dataset more than 20 times.

#### D. Model

LSTM is another modification to RNN, it's also made using the same memory concept that was suggested to record long data sequences before the Gated Recurrent Unit. Thus, GRU is a simplification of LSTM. Activation values are used here in LSTM. There are not only C (candidate), but two cell outputs, a new activation and a new candidate value as well. The value of the new candidate can be determined by Eq. (1)

$$c^{N<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c) \quad (1)$$

LSTM we control the memory cell through 3 different gates Update Gate Eq. (2), Forget Gate Eq.(3), Output Gate Eq. (4).

$$\Gamma_u = \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u) \tag{2}$$

$$\Gamma_f = \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f) \tag{3}$$

$$\Gamma_o = \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o) \tag{4}$$

2 LSTM outputs, the new candidate and a new activation, we would use the previous gates in them. It is shown in Eq. (5) and Eq. (6).

$$c^{<t>} = \Gamma_u \times c^{N^{<t>}} + \Gamma_f \times c^{<t-1>} \tag{5}$$

$$a^{<t>} = \Gamma_o \times c^{<t>} \tag{6}$$

When we link multiple LSTMs together, we can see that if the network learned the parameters of the gates correctly, we could pass the candidate values (red values) from the early

sequence to the very end of the sequence, so that we can model long dependencies with high accuracy. (Fig. 3)

### E. Model hyperparameters

There are various parameters for model construction. The model contains 121,782 pairs of information (text and summary) based on the filtered dataset. These data are separated 80% from the dataset filtered (97,439 data) for training purposes, 12% from the data filtered (14,615 data), and 8% from the data filtered (9,728 data) for the validation of data. For all data the batch size is 32, encoder and decoder have been hidden and learning rate is 0.001 and 0.5. Adam Optimizer is used for optimization.

The other parameters include 4 encoder layers, 4 decoder layers, size of rnn encoder 512, size of rnn decoder 512 with a batch size of 256 and 200 epochs. The keep probability being 0.5 with clip size of 5.

### F. Output

Some of the output using bidirectional LSTM is given in Figs. 4 and 5.

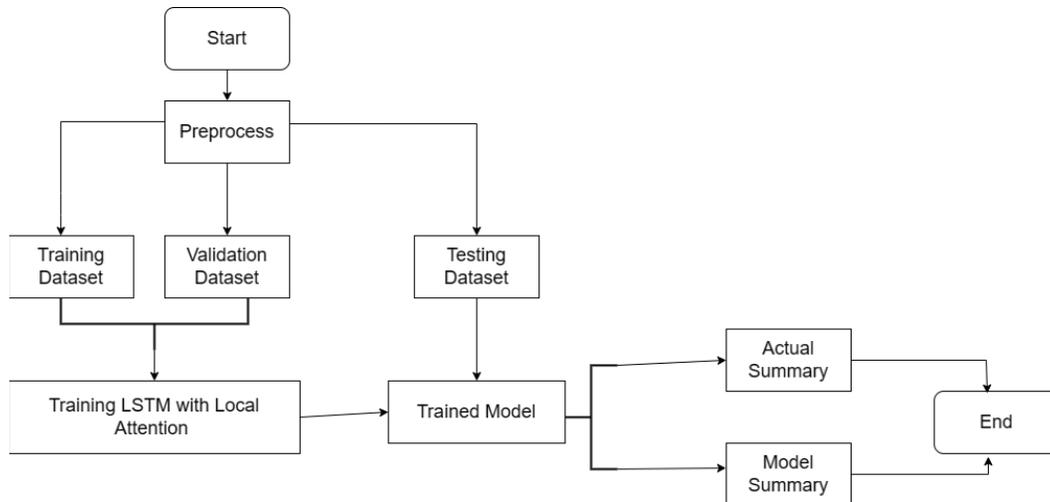


Figure 1. Block diagram of the system.

	Summary	Text
0	Good Quality Dog Food	I have bought several of the Vitality canned d...
1	Not as Advertised	Product arrived labeled as Jumbo Salted Peanut...
2	"Delight" says it all	This is a confection that has been around a fe...
3	Cough Medicine	If you are looking for the secret ingredient i...
4	Great taffy	Great taffy at a great price. There was a wid...

Figure 2. Example Dataset

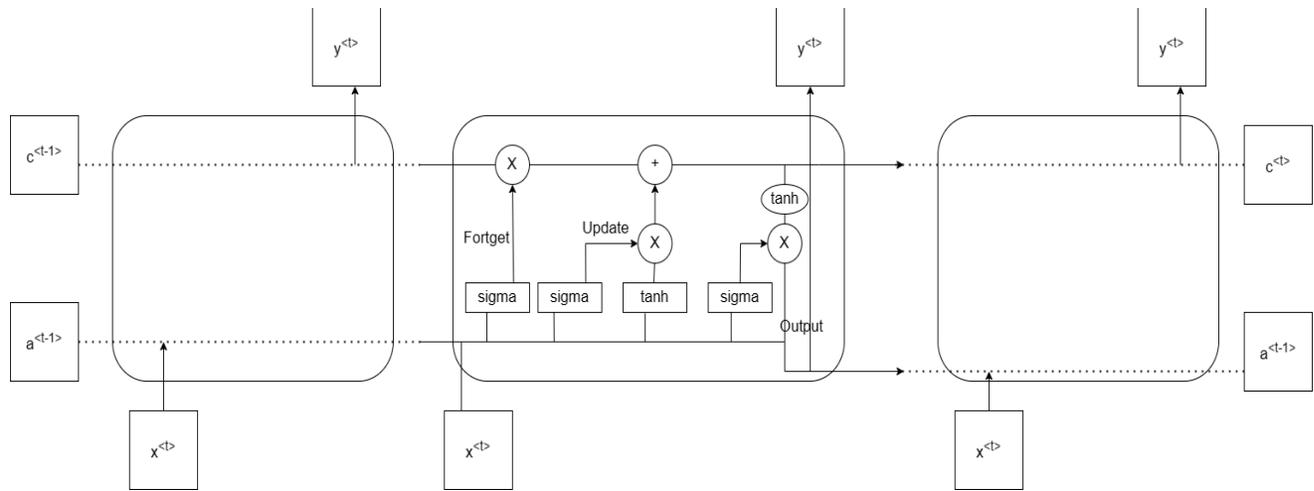


Figure 3. LSTM model structure implementation.

- Review(1): I am not a big fan of cereal but it is nice to have it at the house. Both my husband and my kid likes this. I personally prefer protein shake for breakfast.
- Original summary(1): Family likes it
- My Summary(1): very nice breakfast cereal

- Original Text(2): Good value for a lot of treats! my dog absolutely loves these and it's awesome that they sell them on amazon for half the price of what they would sell it for at a pet store! Will continue buying from here as long as they're available. :)
- Original summary(2): good value!
- My Summary(2): awesome value

- Original Text(3): My baby seems to prefer this formula mixture over the other bigger brands. If you buy it "Subscribe and Save" it's a lot cheaper, which is nice. I breastfed for 9 months until my baby stopped gaining weight so we had to start giving her formula. She's been growing steadily ever since!
- Original summary(3): Seems healthy, and my baby prefers it
- My Summary(3): great formula for baby

- Original Text(4): These are so good, especially when they are still hot. I love the way they still sizzle and pop while eating. Sometimes I sprinkle some hot sauce on mine straight from the microwave or a little parmesan cheese, depending on my mood. However I eat them, they are really good. Remember to shake them well for a minute or two before nuking them. The bags do not puff up as much as microwave popcorn, so don't expect them to. Only cook for the suggested time on the bag.
- Original summary(4): Yummy!
- My Summary(4): good stuff

Figure 4. Summarizer output.

- Original Text(5): IF YOU WANT A HIGH ENERGY, AND GREAT TASTING (AND I MEAN THE VERY BEST)ENERGY DRINKS, THIS IS THE ONE FOR YOU! BEVNET.COM SAYS IT TASTES LIKE MOUNTAIN DEW, AND IT DOES! VOTED BEST TASTING! HAS JIFFY INGREDIENT THAT MAKES THIS DIFFERENT CALLED D-RIBOSE, AND IT IS A SIMPLE SUGAR THAT IS GREAT FOR ATHLETES! NO BAD THINGS HERE. I HAD A BAD BATCH ONCE, BUT IT WAS A FLUKE.
- Original summary(5): THESE ARE THE BEST!
- My Summary(5) : the best energy drink

- Original Text(6): What I really appreciate about these bars is that it really takes time to eat and enjoy them. Highly recommended!
- Original summary(6): Great product!
- My Summary(6) : foolproof boys bars

- Original Text(7): When I was growing up, we ate a fair amount of Spam. Every so often, I revisit those days, by purchasing some Spam and trying to enjoy it. And I think of the old Monty Python segment focusing on Spam. . . .

It's not so easy for me to enjoy! Plain Spam, even cooked, doesn't do it for me. I start out with good intentions of eating it as part of a meal and normally throw quite a bit of it out. So, this time around, I tried something different--a fried Spam sandwich.

I cut up and fried some shallots, slathered a couple pieces of bread with Dijon mustard and catsup, and melted some American cheese over the Spam. And it worked! This time, I successfully fixed something that didn't feature the Spam taste. The result was an enjoyable sandwich.

So, instead of throwing out all the Spam (as has happened in the past), I'm going to use my imagination this time in preparing this for a meal. In the past, this would have rated 1 to 2 stars, this time around, I'll give it three.

- Original summary(7): Back to the future with Spam
- Response Words(7): not the best

Figure 5. Summarizer output.

### 3.2. Model using pointer generator method

The seq2seq+care framework is the basis for our model. In the word embedding layer, the source text is not encoded after the word segmentation; the part of speech is also encoded, and the two matrices are used together for model input. Besides the two-way LSTM that can be used for learning. CNN, which is capable of extracting n-gram and part-of- speech features, are also added to the source text sentence level features. Then we combine the above learning matrices into the context vector. The OOV and self-repetition problem can be solved by the decoder. Effectively generate by adding a network of points and generators. This is shown in the Fig.6.

### A. Dataset

The dataset used to train the model is the “CNN / Daily Mail” newspaper readings. It contains 300,00 pairs of news articles (781 tokens) and multi-sentence summaries (56 tokens). Built to contain multiple summaries for the same story, the data is supplied to the model via a script which transforms it into a binary chunk file and then is supplied to the model. The dataset is split into (Training: Developing:Testing) 92:4.2:3.8. It is shown in Fig.7.

This process then is followed by the existing process of Preprocessing, Filtration as given under heading 3.1.B and 3.1.C

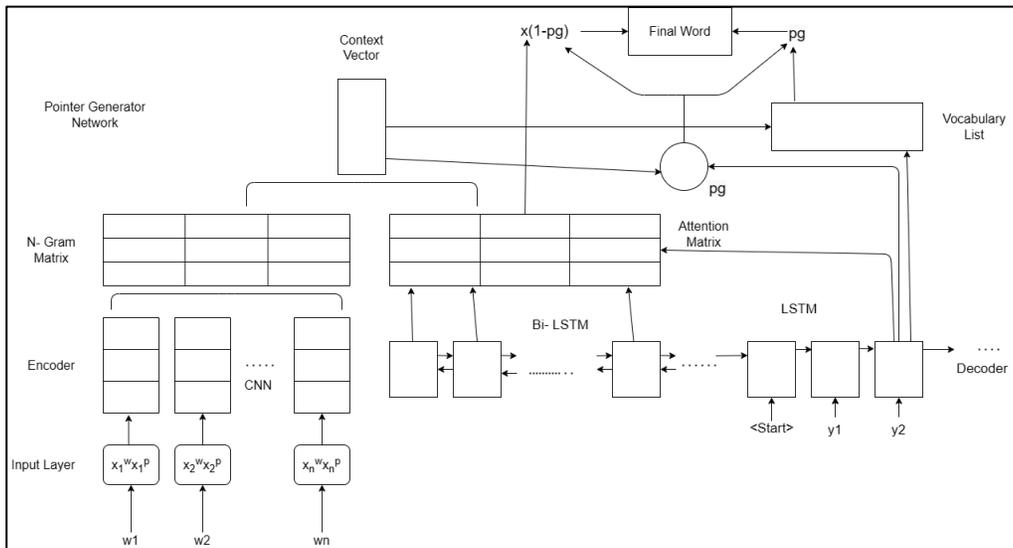


Figure 6. Pointer generator model.

```
<example>
<article>england 's campaign at the european under 21 championship looks set to be dominated over whether harry kane is included - but gareth southgate may not need him . in front of over 30,000 at the riverside stadium , england came from behind to defeat germany 3-2 and record a victory against not just old rivals but the pre-tournament favourites in the czech republic this summer . southgate is still expected to do everything he can to include the tottenham striker , but the win over germany shows that the young lions can go toe-to-toe with the cream of europe . here sportsmail looks at england 's three stand-out stars from the win in middlesbrough . james ward-provse . the southampton star has already become an established member of the first team at st mary 's and is set to be a key member of southgate 's side in the czech republic . the 20-year-old scored the winner with a crisp low effort after __ghosting__ into the penalty box to find space . but his attacking play extends to set-pieces . germany 's julian __korb__ reaches out in vain as james ward-provse fires home england 's winning goal . the under 21 captain celebrates his late strike following a fine midfield performance . in ward-provse , england have potentially their best __dead-ball__ specialist since david beckham with his passing range and accurate shooting already a highlight of his game . ward-provse was captain for the germany victory and his rapid progress on the south coast in the last two years suggests he is worth keeping an eye on . carl jenkinson . named the official man of the match , jenkinson put in a close to perfect performance - albeit he was slightly at fault for germany 's opening goal after allowing the cross for philipp __hofmann__ to fire home . but after a slow start to each half , the west ham full-back , on loan from arsenal , imposed himself down the right-hand side . carl jenkinson runs the ball out of defence for england as germany substitute nico schulz looks on . jenkinson put in a man of the match display having assisted two of england 's three goals . it was the 23-year-old 's cross for jesse lingard which helped draw the three lions level in the first half , while he also provided the pass for ward-provse to tuck home the winning goal . with england 's senior team lacking depth in the right-back area , roy hodgson could do worse than pick jenkinson . because if he does n't , then finland will as he can still play for the scandinavian team . nathan redmond . relished the battle with germany full-back christian gunter , often managing to pull away from his marker to lead an attack down the right wing . redmond is not all about pace though as he often tried to cut inside to mix up his approach , with his passing in the final third managing to catch the german back four flat on a number of occasions . nathan redmond rifles home england 's second equaliser across goal into the bottom corner . redmond celebrates after drawing england level late in the game following an impressive display on the right . it was a much improved version of the winger who struggled in a poor norwich side in the premier league last season , but having struck the equaliser against a talented germany team he looks to be improving . now a veteran of the under 21 set-up , it will be difficult for southgate to leave him out of the side in the summer to help make space for kane as he has often impressed on the right wing for the young lions .</article>
```

Figure 7. Example dataset

### B. Model

The model contains 2 input parts. 1) word embedding vector  
2) speech vector defined by Eq. (7) and Eq. (8) respectively.

$$x^w = \{x_1^w, x_2^w, \dots, x_n^w\} \tag{7}$$

$$x^p = \{x_1^p, x_2^p, \dots, x_n^p\} \tag{8}$$

The basic model is the model seq2seq+care. We use two-way LSTM for the encoder. In order to create a sequence of hidden encoder state, the word embedding vectors are fed to the encoder one by one. Because we always use a two-way LSTM.

Step takes both the last and the next hidden state and we are connecting it to the hidden status of the encoder  $h_t^*$ . Eq.(9)

$$p_{gen}(s_i=1) = \sigma(v^s * (W_h^s * h_t + W_e^s * E[o_{i-1}] + W_c^s * c_i + b^s)) \tag{9}$$

$h_i$ =hidden state of decoder;  $E[o_{i-1}]$  = time step of decoder;  $c_i$ = attention weighted context vector.  $W_h^s, W_e^s, W_c^s, b^s, v^s$ , are learning parameters

We introduce the network of pointer generators in our model to solve the problem of OOV. It can copy a word from the source text or create one using its pointing mechanism from its vocabulary. The probability is defined by the context vector  $cont$ , the decoder status  $st$  and the decoder input  $xt$  for the decoder time step. Eq.(10)

$$p_{gen} = \sigma(W_h^T * h_t^* + W_s^T * s_t + W_x^T * x_t + b_{ptr}) \tag{10}$$

$s_i$ =decoder state;  $x_i$ = decoder input;  $h_t^*$ =context vector.  
 $W_h^T, W_s^T, W_x^T, b_{ptr}$  are learning parameters.

### C. Model hyperparameters

For 100,000 iterations we have trained the pointer generator network using a 0.15 learning rate. We used a size 8 batch and a size 4 beam. The hidden state was 256 and the embedding was 128. The vocabulary had been 50,000. The training took 22 hours.

### D. Output

Some of the output using Pointer Generator with bidirectional LSTM is as follows: (Fig. 8; Fig. 9; Fig. 10; Fig. 11)

```
<article>england 's campaign at the european under 21 championship looks set to be dominated over whether harry kane is included - but gareth southgate may not need him . in front of over 30,000 at the riverside stadium , england came from behind to defeat germany 3-2 and record a victory against not just old rivals but the pre-tournament favourites in the czech republic this summer . southgate is still expected to do everything he can to include the tottenham striker , but the win over germany shows that the young lions can go toe-to-toe with the cream of europe . here sportsmail looks at england 's three stand-out stars from the win in middlesbrough . james ward-provse . the southampton star has already become an established member of the first team at st mary 's and is set to be a key member of southgate 's side in the czech republic . the 20-year-old scored the winner with a crisp low effort after __ghosting__ into the penalty box to find space . but his attacking play extends to set-pieces . germany 's julian __korb__ reaches out in vain as james ward-provse fires home england 's winning goal . the under 21 captain celebrates his late strike following a fine midfield performance . in ward-provse , england have potentially their best __dead-ball__ specialist since david beckham with his passing range and accurate shooting already a highlight of his game . ward-provse was captain for the germany victory and his rapid progress on the south coast in the last two years suggests he is worth keeping an eye on . carl jenkinson . named the official man of the match , jenkinson put in a close to perfect performance - albeit he was slightly at fault for germany 's opening goal after allowing the cross for philipp
```

Figure 8. Article (First part).

```
__hofmann__ to fire home . but after a slow start to each half , the west ham full-back , on loan from arsenal , imposed himself down the right-hand side . carl jenkinson runs the ball out of defence for england as germany substitute nico schulz looks on . jenkinson put in a man of the match display having assisted two of england 's three goals . it was the 23-year-old 's cross for jesse lingard which helped draw the three lions level in the first half , while he also provided the pass for ward-provse to tuck home the winning goal . with england 's senior team lacking depth in the right-back area , roy hodgson could do worse than pick jenkinson . because if he does n't , then finland will as he can still play for the scandinavian team . nathan redmond . relished the battle with germany full-back christian gunter , often managing to pull away from his marker to lead an attack down the right wing . redmond is not all about pace though as he often tried to cut inside to mix up his approach , with his passing in the final third managing to catch the german back four flat on a number of occasions . nathan redmond rifles home england 's second equaliser across goal into the bottom corner . redmond celebrates after drawing england level late in the game following an impressive display on the right . it was a much improved version of the winger who struggled in a poor norwich side in the premier league last season , but having struck the equaliser against a talented germany team he looks to be improving . now a veteran of the under 21 set-up , it will be difficult for southgate to leave him out of the side in the summer to help make space for kane as he has often impressed on the right wing for the young lions .</article>
```

Figure 9. Article (Second part).

```
<reference>england recorded impressive 3-2 win over germany in under 21 friendly . young lions twice came from behind to seal late win at riverside stadium . james ward-provse , carl jenkinson and nathan redmond among stars . gareth southgate 's side look in strong shape heading to the european under 21 championship in czech republic this summer .</reference>
<summary>the southampton star has already become an established member of the first team at st mary 's and is set to be a key member of southgate 's side in the czech republic . the southampton star has already become an established member of the first team at st mary 's and is set to be a key member of southgate 's side in the czech republic .</summary>
```

Figure 10. Reference and Summary.

```
<eval>
<BLEU score="1.210743289604319"/>
<ROUGE_1 score="0.1212121212121212"/>
<ROUGE_2 score="0.0625"/>
<ROUGE_L score="0.1212121212121212"/>
<ROUGE_be score="0"/>
</eval>
```

Figure 11. Evaluation of summary.

## 4. Result

The result of the models used in this paper is given by graphs on various comparison parameters like ROUGE-1, ROUGE-2, ROUGE-L (Fig.12); Accuracy (Fig. 13) and Loss Graphs (Fig.14).

The values of the above graphs are shown in the tables as below (Table 1, Table 2, Table 3).

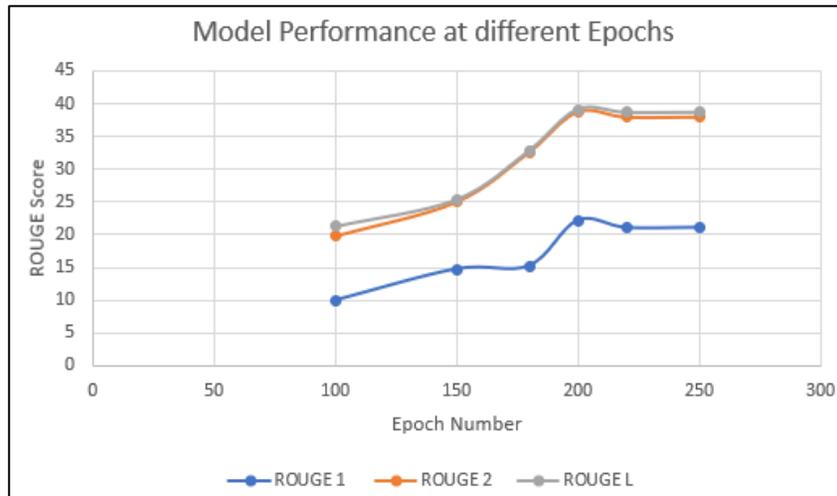


Figure 12. ROUGE-1, ROUGE-2, ROUGE-L performance of both models.

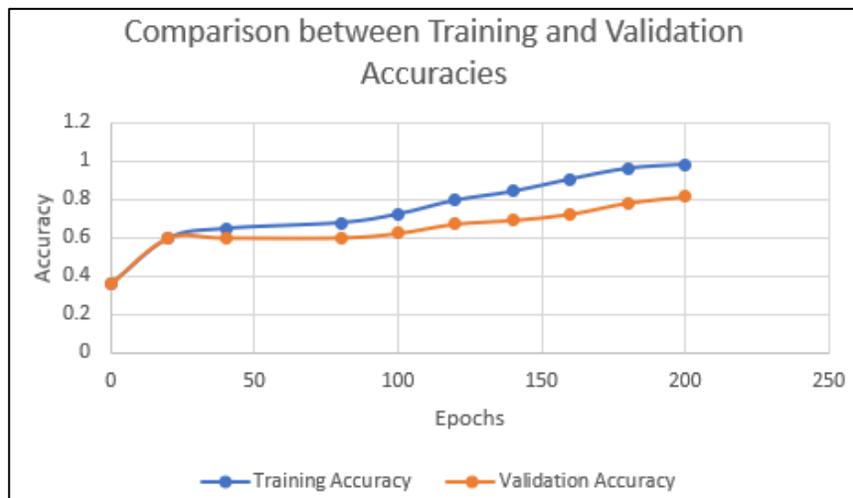


Figure 13. Accuracy of both the models

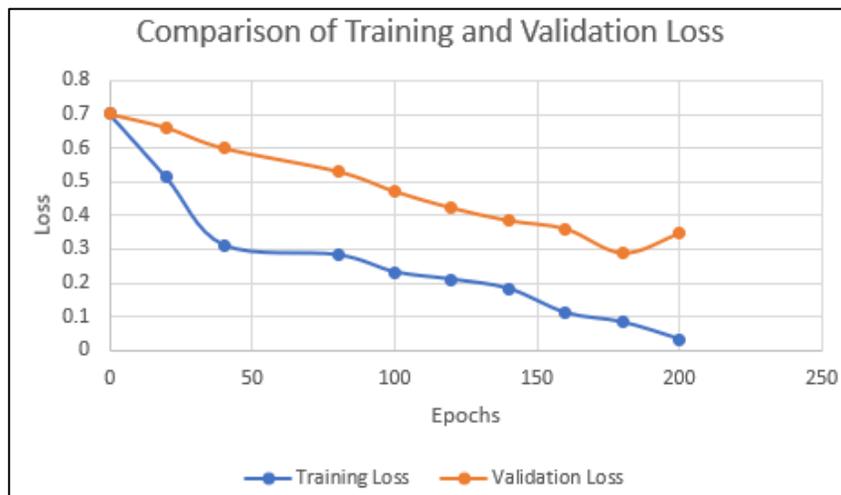


Figure 14. Loss of both models.

Table 1. Particle size and power consumption.

Epoch No.	ROUGE 1	ROUGE 2	ROUGE L
100	10	19.85	21.25
150	14.75	25.05	25.35
180	15.25	32.65	32.85
200	22.16	38.76	39.12
220	21.05	37.96	38.65
250	21.058	37.963	38.653

Table 2. Comparison between training and validation accuracy of both models.

Epoch No.	Training Accuracy	Validation Accuracy
0	0.356	0.356
20	0.5968	0.5969
40	0.6485	0.5996
80	0.6785	0.6005
100	0.7235	0.6235
120	0.7956	0.6723
140	0.8413	0.6914
160	0.9059	0.7238
180	0.9613	0.7826
200	0.9817	0.8145

Table 3. Comparison between training and validation loss of both models.

Epoch No.	Training Loss	Validation Loss
0	0.7029	0.7029
20	0.5126	0.6621
40	0.3128	0.6015
80	0.2836	0.5325
100	0.2324	0.4729
120	0.2103	0.4237
140	0.1824	0.3862
160	0.1102	0.3614
180	0.0823	0.2897

## 5. Discussion

An important question in Natural Language Processing is that- Why do we employ a sophisticated network structure rather than a basic neural network? This is because It is critical for the network to comprehend the word itself, rather than linking the term to a specific place.

Here we utilise Bidirectional networks which are adjustments made to the standard RNN network to allow it to adapt to a critical necessity in NLP situations. In NLP, sometimes understanding a word requires referring not just to

the preceding word, but also to the next word. To obtain even better results, we may stack numerous RNNs on top of each other, but keep in mind that they function with time.

The text summarization model can be viewed as a condition model for NLP which implies that to produce a summary of a given input which results in the output be based and conditioned on the input. But the outputs will be variable and form a large pool of outputs so to solve the problem we can generate one word during a period of time, then next word and so on but this method increases time and space complexity by many folds. So we can solve this problem by

applying Beam Search which utilises all probabilities of all words in the dataset.

When we humans summarise material, we glance at a few words at a time, not the entire text to summarise at one time; this is what we are attempting to teach our model. Here we create a seq2seq encoder decoder which utilises the interface of context vector.

This research can be trained to take on new words without creating dictionaries and datasets of available vocabulary and apply new language models used in modern English and languages with help of Reinforcement Learning.

## 6. Conclusions

This paper presented the English text summary with the encoding and decoding of LSTM as a good approach. But no accurate 100% value or prediction is given to all machines. Our summarizer provides a very precise summary, as does a machine. Few reviewed, but it offers excellent performance and a fluent synthesis for other reviews, cannot be predicted accurately. After all, with the reduction of training loss we can successfully create a comprehensible, fluid and short summary.

The global attention model produces a better ROUGE-1 where more words are generated in the current synopsis. However, ROUGE-2 gives more local attention, where more pairs of words are generated in the actual summary, since the local attention mechanism takes account of the sub-set of entry words instead of the whole entry word. The dataset contains many symbols and unknown because it is written using informal words. Sentences which are not in the word data set for embedding. The ROUGE score is therefore not higher than the usual English text score. All parameters can be reset for both models, give higher scores. Some methods to improve both models can be developed to change the dataset to any other article text. Reconstruction of the model with more optimal parameters or handling of an OOV in preprocessing data instead of reviewing text.

We conclude from our analysis of the summary subtask of the op-ed article that Up-to - date recurring models are successful in traditional news articles, not widespread Well to subtasks that are harder. Many work related to the inclusion of extractive measures to cut off An abstractive summary model before training. Input. However, we believe that masking the input in advance, in the case of harder unstructured components such as op-eds, would further degrade performance and dilute the simplicity of the resuming pipeline.

## Conflict of interest

The authors have no conflict of interest to declare.

## Acknowledgments

We would like to acknowledge all those who have helped with the completion of this project.

First, I, Saroj Anand Tripathy would like to extend our gratefulness to Professor Sharmila A, for her valuable advice, continual support, suggestions and patience during the course of project.

Our special thanks are also extended to our friends for their academic, moral support and furthermore their helpful assistance during the duration of our project. Finally, we would like to thank Vellore Institute of Technology University, Vellore for providing support to complete the project in successful manner.

## Financing

The authors received no specific funding for this work.

## References

- Afsharizadeh, M., Ebrahimpour-Komleh, H., & Bagheri, A. (2018). Query-oriented text summarization using sentence extraction technique. 2018 4th *International Conference on Web Research (ICWR)*, pp.128-132. <https://doi.org/10.1109/icwr.2018.8387248>
- Alfarra, M. R., Alfarra, A. M., Alattar, J. M. (2019). Graph-based fuzzy logic for extractive text summarization (GFLES). 2019 *International Conference on Promising Electronic Technologies (ICPET)*, pp. 96-101. <https://doi.org/10.1109/icpet.2019.00025>
- Chen, J., & Zhuge, H. (2018). Extractive text-image summarization using multi-modal RNN. 2018 14th *International Conference on Semantics, Knowledge and Grids (SKG)*, pp.245-248. <https://doi.org/10.1109/skg.2018.00033>
- Devasena, C. L., & Hemalatha, M. (2012). *Automatic text categorization and summarization using rule reduction*. In *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM-2012)* (pp. 594-598). IEEE.

- Geng, Z., Zhang, J., Li, X., Du, J., & Liu, Z. (2010). Post-Processing of Automatic Text Summarization for Domain-Specific Documents. In *2010 International Conference on Communications and Mobile Computing* (Vol. 1, pp. 387-391). IEEE.  
<https://doi.org/10.1109/CMC.2010.143>
- Gunawan, D., Harahap, S. H., & Rahmat, R. F. (2019). Multi-document summarization by using textrank and maximal marginal relevance for text in Bahasa Indonesia. In *2019 International conference on ICT for smart society (ICISS)* (Vol. 7, pp. 1-5). IEEE.  
<https://doi.org/10.1109/ICISS48059.2019.8969785>
- Hanunggul, P. M., & Suyanto, S. (2019). The impact of local attention in LSTM for abstractive text summarization. *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pp.54-57.  
<https://doi.org/10.1109/isriti48646.2019.9034616>
- Hark, C., Uckan, T., Seyyarer, E., & Karci, A. (2018). Graph-based suggestion for text summarization. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)* (pp. 1-6).  
<https://doi.org/10.1109/idap.2018.8620738>
- Iwasaki, Y., Yamashita, A., Konno, Y., & Matsubayashi, K. (2019, November). Japanese abstractive text summarization using BERT. In *2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI)* (pp. 1-5). IEEE.  
<https://doi.org/10.1109/taai48200.2019.8959920>
- Jiang, X. Y., Fan, X. Z., Wang, Z. F., & Jia, K. L. (2009). Improving the performance of text categorization using automatic summarization. In *2009 International Conference on Computer Modeling and Simulation* (pp. 347-351). IEEE.  
<https://doi.org/10.1109/iccms.2009.29>
- Jo, T. (2017). K nearest neighbor for text summarization using feature similarity. In *2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)* (pp. 1-5). IEEE.  
<https://doi.org/10.1109/iccccee.2017.7866705>
- Krishnaveni, P., & Balasundaram, S. R. (2017). Automatic text summarization by local scoring and ranking for improving coherence. In *2017 international conference on computing methodologies and communication (ICCMC)* (pp. 59-64). IEEE.  
<https://doi.org/10.1109/iccmc.2017.8282539>
- Lukyamuzi, A., Ngubiri, J., & Okori, W. (2019). Topic Based Machine Learning Summarizer. In *2019 IEEE International Smart Cities Conference (ISC2)* (pp. 288-291). IEEE.  
<https://doi.org/10.1109/isc246665.2019.9071737>
- Madhuri, J. N., & Kumar, R. G. (2019). Extractive text summarization using sentence ranking. In *2019 international conference on data science and communication (IconDSC)* (pp. 1-3). IEEE.  
<https://doi.org/10.1109/icondsc.2019.8817040>
- Mishra, A. R., Panchal, V. K., & Kumar, P. (2019). Extractive Text Summarization-An effective approach to extract information from Text. In *2019 International Conference on contemporary Computing and Informatics (IC3I)* (pp. 252-255). IEEE.  
<https://doi.org/10.1109/ic3i46837.2019.9055636>
- Masum, A. K. M., Abujar, S., Talukder, M. A. I., Rabby, A. S. A., & Hossain, S. A. (2019). Abstractive method of text summarization with sequence to sequence RNNs. In *2019 10th international conference on computing, communication and networking technologies (ICCCNT)* (pp. 1-5). IEEE.  
<https://doi.org/10.1109/icccnt45670.2019.8944620>
- Pal, A. R., & Saha, D. (2014). An approach to automatic text summarization using WordNet. In *2014 IEEE international advance computing conference (IACC)* (pp. 1169-1173). IEEE.  
<https://doi.org/10.1109/iadcc.2014.6779492>
- Pattnaik, S., & Nayak, A. K. (2019). Summarization of odia text document using cosine similarity and clustering. In *2019 International Conference on Applied Machine Learning (ICAML)* (pp. 143-146). IEEE.  
<https://doi.org/10.1109/icaml48257.2019.00035>
- Prakash, C., & Shukla, A. (2014). Human aided text summarizer" saar" using reinforcement learning. In *2014 International Conference on Soft Computing and Machine Intelligence* (pp. 83-87). IEEE.  
<https://doi.org/10.1109/iscmi.2014.22>
- Rashidghalam, H., Taherkhani, M., & Mahmoudi, F. (2016). Text summarization using concept graph and BabelNet knowledge base. In *2016 Artificial Intelligence and Robotics (IRANOPEN)* (pp. 115-119). IEEE.  
<https://doi.org/10.1109/rios.2016.7529500>

Reategui, E., Klemann, M., & Finco, M. D. (2012). Using a text mining tool to support text summarization. In *2012 IEEE 12th International Conference on Advanced Learning Technologies* (pp. 607-609). IEEE.

<https://doi.org/10.1109/icalt.2012.51>

Ren, S., & Zhang, Z. (2019). Pointer-Generator Abstractive Text Summarization Model with Part of Speech Features. In *2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)* (pp. 1-4). IEEE.

<https://doi.org/10.1109/icsess47205.2019.9040715>

Ren, S., & Guo, K. (2019). Text Summarization Model of Combining Global Gated Unit and Copy Mechanism. In *2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)* (pp. 390-393). IEEE.

<https://doi.org/10.1109/icsess47205.2019.9040794>

Sherry, & Bhatia, P. (2015). Multilingual text summarization with UNL. In *2015 International Conference on Advances in Computer Engineering and Applications* (pp. 740-745). IEEE.

<https://doi.org/10.1109/icacea.2015.7164787>

Zenkert, J., Klahold, A., & Fathi, M. (2018). Towards extractive text summarization using multidimensional knowledge representation. In *2018 IEEE International Conference on Electro/Information Technology (EIT)* (pp. 0826-0831). IEEE.

<https://doi.org/10.1109/eit.2018.8500186>

Zhang, P. Y., & Li, C. H. (2009). Automatic text summarization based on sentences clustering and extraction. In *2009 2nd IEEE international conference on computer science and information technology* (pp. 167-170). IEEE.

<https://doi.org/10.1109/iccsit.2009.5234971>