ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Enhancing Text Summarization using Hybrid LSTM-GRU with Lingual Significance Relation-based Attention Mechanism Model

PL. Prabha[1] and M. Parvathy[2]

ABSTRACT

Text summarization is the process of summarizing the information of a large text into short, crisp, and concise text to analyze and extract the most imperative information from the given text. Therefore, different AI based techniques are used for summarizing the text. In order to achieve this, various AI techniques have been incorporated into the existing works. However, the prevailing methods lagged in delivering accurate text. Therefore, the proposed work employs a Hybrid LSTM-GRU (Long Short Term Model –Gated Recurrent Unit) model with M-AM (Modified - Attention Mechanism). The dataset incorporated in the proposed model is amazon fine food review. Different pre-processing techniques remove unwanted and irrelevant text, such as tokenization, text cleaning, stop word removal, and stemming and lemmatization. The Proposed model employs hybrid LSTM-GRU with M-AM as it delivers faster and employs less memory consumption. Along with it, it has the potential to capture long-term dependencies as well. Further, M-AM incorporates lexical sequence measure and sentence context weight for delivering an effective model for text summarization. Therefore, the major contribution of the proposed work involves summarizing the text into a crisp and brief format for easy understanding. Finally, the performance of the proposed model is evaluated using different ROUGE, accuracy, and loss, in which ROUGE metrics obtained by the proposed model is 55.5.

Article information:

Keywords: Text Summarization, Natural Language Processing, Long Short Term Memory, Gated Recurrent Unit, ROUGE

Article history:
Received: August 18, 2023
Revised: November 30, 2023
Accepted: February 1, 2024
Published: April 13, 2024 (Online)

DOI: 10.37936/ecti-cit.2024182.253853

1. INTRODUCTION

Text summarization delivers a brief and crisp summary by apprehending significant information and widespread meaning [1]. Text summarization is accomplished by NLP approaches by employing different algorithms for achieving better outcomes. Due to the exponential growth of the content and the need to extract key information effectively, text summarization has attained significant attention in the last few years [2]. Text summarization has come into existence in the year 1958 [3]. The methods for text summarization include extractive and abstractive summarization. Text summarization is performed using various techniques, such as attaining the occurrence or frequency in a sentence or the existence of a word or sentence in a specific place in the sentence or paragraph. While summarizing the text, there are possibilities of various unwanted or irrelevant data, which can be removed during the pre-processing stage using various pre-processing techniques, such as removal of stop words [4], lemmatization, stemming [5], tokenization [6], removal of punctuations [7] and other such techniques.

According to various studies, LSTM (Long Short Term Memory) is predominantly used for analyzing the text, speech, and voice sequence-to-sequence (Seq2Seq) data analyzing process. Likewise, Kovacevic and Keco [8] used Amazon's database of food and product review datasets have been used for text summarization. The encoder-decoder architecture, with stacked LSTMs in the encoder phase and an attention layer mechanism, has been used for text summarization. In addition, bidirectional LSTMs were used

[1,2] The authors are the Department of Computer Science and Engineering, Sethu Institute of Technology Madurai, India, E-mail: prabha_pl2003@yahoo.co.in, prabhasethuraman18@gmail.com and parvathydurai2002@gmail.com
[1] Corresponding author: prabha_pl2003@yahoo.co.in and prabhasethuraman18@gmail.com

instead of the unidirectional approach. The combination of LSTM and stacking results in increased sequence representation and quality in summarization. The produced outcome of the study is generally compatible with original summaries that convey the same intellectual meaning. In recent days, an enormous amount of textual information has become available throughout the internet. The information has been scattered in many places in many forms. The process of summarizing and generating concise textual content without altering the original meaning is a challenging task [9]. Therefore, methods like sequence-to-sequence deep neural network models such as Long Short Term Memory (LSTM) or Gated Recurrent Unit (GRU) were used by Nguyen [10] to handle the textual formats. The process consists of two steps: encoding the input and generating the summary. The LSTM model has lost some input information during text encoding and produces inappropriate results, which leads to muted output. The Recurrent Residual Attention mechanism has tested Amazon's dataset and outperformed previous methods.

An enhanced approach has been combined with the Deep Learning technique for text summarization. Some feature extraction models are listed below: k-means, affinity propagation, DBSCAN, and PageRank. The deep Learning model is supposed to propose a system for unsupervised abstractive summarization. The sample dataset was collected from Amazon's product reviews, consisting of 142.8 million reviews from May 1996 to July 2014 [11]. Practically difficult to evaluate and summarise when it comes to unsupervised data. To achieve the expected performance, some suitable metrics are used, like ROUGE-1, semantic similarity, sentiment accuracy, attribute match, and content preservation, to measure the summary's quality. A collective approach has been created for text summarization, which examined the raw data and produced robust extractive and abstract summarization. RNN and LSTM models are used to create the extractive summarization. The output created from this architecture is taken as an input source for abstractive summarization, and the Pointer Generator Network is used for implementation. The standard CNN/daily mail dataset has been examined for experimental purposes. The results are evaluated using ROUGE scores [6]. Pieta demonstrated [12] the Pytorch-nuevo ML technique for Deep Learning neural network computation. The framework has been capable of solving NLP tasks, which include text summarization based on RNN, LSTM, and seq2seq.

For examining the task, Amazon's and IMDB Movie reviews are used in ROUGE-N, BLEU, and F1 algorithms. Muthiah used [13] logical and human-readable textual summaries of the online product reviews collected over a period of time. The collected samples are given to the abstractive and extractive approaches to attain more concise results. The Combined Extractive Abstractive Text Summarization (CEATS) model and sequence to sequence encoder and decoder model made for RNN LSTM networks to achieve the best results. Likewise, an attention mechanism was used [14], which is termed Neural Machine Translation (NMT). In general, two classes of attention are discussed. Local and Global attention is known as attention classes. The study focuses on employing local attention in the Long Short-Term Memory (LSTM) model to generate Abstractive text Summarization (ATS) for the Amazon Fine Food review dataset used for evaluation. ROUGE-1 and ROUGE-2 performance are compared, and the desired output is achieved

Though existing studies delivered better results for text summarization, they still lagged in delivering fast and accurate outcomes in accordance with text summarization due to ineffective algorithms. Therefore, the proposed study employed a hybrid LSTM-GRU with the M-AM model for text summarization, as the proposed model has the potential to deliver fast, concise, crisp, and accurate outcomes for text summarization and can summarize the original text without changing the actual meaning of it. Further, M-AM comprises 2 techniques, lexical sequence measure and sentence context weights, for producing an effective and efficient model for text summarization. Objectives of the research include,

- To pre-process the data using text cleaning, stop word removal, Tokenization, stemming, and lemmatization.
- To summarize the text from large paragraph to small using hybrid LSTM-GRU with M-AM for fast and accurate text summarization
- To assess the efficiency of the proposed model using ROUGE, accuracy, and loss metrics.

1.1 Paper Organization

Section II deals with conventional methods done on a similar domain with diverse methods, as shown in Further, Section III represents the methodology executed in the projected system. The results and outcomes accomplished by the projected method are shown in Section IV. Finally, the conclusion and future work of the projected system is shown in Section V.

2. LITERATURE REVIEW

Various existing works with different algorithms for text summarization are mentioned in the subsequent section.

Humans can summarize complex and protracted documents in a simple and brief format. However, when it comes to processing and summarizing huge volumes of documents within a fraction of a second, humans do not possess the capability to do so, whereas machines can [15]. Therefore, a text summarization model was focused on by Atanda [16], which aided in summarizing huge volumes of customer reviews extracted from the Amazon dataset. The process was carried out using the TextRank algorithm, and the values obtained were fed to the LSTM algorithm to produce the summary of the text. Likewise, the amazon fine food review dataset was used by Masum [17] for summarizing the text from a large document. To achieve text summarization, Bi-RNN with LSTM has been used in the encoding layer, and in the decoding layer, the attention mechanism has been used. The objective of the model was to increase the efficiency of the model and aid in reducing the training loss of the sequence-to-sequence model to create an abstractive text summarizer.

Correspondingly, a readable, crisp, and informative summary of the reviews was primarily focused by Debnath [18], as these reviews assisted in making a purchase decision for the customer. Therefore, the LSTM model has been used, in which the first stage dealt with data-pre-processing of the model aided in creating an organized representation of the text, and the second stage of the model was based on the attention mechanism of the LSTM model, which was trained, tested, and validated. Likewise, Manore [19] used Bi-LSTM, LSTM, and LSTM with an attention mechanism to summarize text using an Amazon review dataset. Initially, the dataset was pre-processed using various pre-processing techniques. Further, the dataset was transformed, and features were selected and transformed using vector representation; then, DL models were employed for the summarization of text, and eventually, the performance of the model was assessed using the ROUGE metric. However, there have been a few limitations of employing BiLSTM, LSTM, and LSTM with AM, which include insufficient computational power on the dataset.

Extensive information has been summarized into precise information using a text summarizer, Bi-LSTM, and PGM (Pointer Generator Mode) [20]. The LSTM model employed has been trained and tested on the AFFR dataset. The model aimed to deliver a dependable model for summarizing the data in the dataset. Significant sentences were fetched by eliminating the unwanted and irrelevant sentences. Different parameters have been used to evaluate the efficiency of the existing model. Different parameters, such as Cosine similarity, ROUGE 1, ROUGE L, and ROUGE 2, were employed to evaluate the model's efficacy [20]. Data has been reachable and obtainable all over the world in humongous numbers. However, it is quite challenging to read long texts as they can become tedious. Hence, Singh [21] utilized the D-NN model for text summarization as it comprised of sequence to-sequence encoder model, which was the LSTM model. Data preprocessing techniques were employed for generating precise output, including removing duplicate values, short words, special characters, punctuations, symbols, etc. The different approaches were handled in the D-NN model by employing sequence to sequence using attention mechanism (AM) and sequence to sequence using encoder–decoder and finally, summarization of texts by employing GPT-2, NLTK, and BERT model. ROUGE metrics such as ROUGE-1, ROUGE 2, and ROUGE -3 have been used by the model for evaluating the accuracy.

The study of Divya [22] focused on the LSTM model, which comes under a Type of RNN model. The pre-processing of the data involves converting everything to lowercase, Removing ', eliminating any special characters and punctuations, and removing stop and short words. The LSTM model was employed for text summarization as LSTM employed backpropagation to train the model. Due to this, the LSTM model could be employed for text summarization techniques. Similarly, different methods like LSTM, BART, Pegasus, and the BART-Large model have been utilized by Mercan [23] for summarizing the text. News summary dataset, amazon fine food review dataset, and so on were used, and the outcome of the study has shown that BART-Large was identified to provide better performance than the other existing ones. Likewise, Sanjabi [24] fixated on using DL architecture in NLP as it aids in better text summarization. It has been believed that DL architecture, aided in NLP, has resulted in better outcomes for text summarization. In most cases, a better outcome has been delivered by adding an Attention mechanism to RNN. Different pre-processing steps have been employed in the study to reduce the model's noise. Therefore, pre-processing techniques like removingl of stop words, duplicate words, bad characteand rs, and white spaces were conpart of thepart of the pre-processing process. [24].

Similarly, summarizing online reviews in the text was carried out by Sheela [25] by incorporating the RNN-LSTM model along with the RVA and CM (copy mechanism) to generate a summary for a particular text. The RNN with RVA process has been trained via FF-NN with encoder-decoder to solve the summarization [25]. Another DL method incorporated in the study by Boumahdi [26], where CNN and auto-encoder were used as the primary objective of the study, was to generate the best and most precise summary for Amazon food review datasets. Further, the paragraph has been rebuilt using the ROUGE metric [26].

Abstractive and Extractive summarization

Text summarization can be accompanied by employing 2 methods, which include abstractive summarization and extractive summarization [27]. However, very less comparison has been made for these techniques. Hence, Bhargav [28] has emphasized combining sequence two sequence decoder with attention and extractive model usually comprised of KNN algorithm, BERT model for Amazon fine food review dataset. The pre-processing steps involved in the study have been the removal of HTML tags, stop-words, special characters, and so on. Different models were compared, in which the BERT model performed better than the existing KNN model during the training as well as the testing process by using various evaluation metrics such as ROUGE-1, soft cosine similarity, count vectorizer, and Tf-idf. Future work of the study emphasized distinguishing different models for text summarization. Therefore, effective models aided in delivering better text summarization and finding better text quality [28].

2.1 Problem Identification

From the assessment of the above-existing works, core concerns are emphasized as explored below,

- The value of ROUGE obtained in the projected study is low, which makes the model inefficient for text summarization [18].
- The existing study is slow and inaccurate for text summarization, which makes the process of text summarization lousy [28].

3. PROPOSED METHODOLOGY

Text summarization is the process of generating short, crisp, concise, and, most importantly, accurate summaries from a long and huge document. However, problems can arise when summarizing long or inappropriate paragraphs into crisp and concise ones, which include inaccurate and irrelevant text. Therefore, various DL techniques were employed for text summarization, however, prevailing models are not capable enough to summarize the text in a fast and accurate manner, which can be overcome by using the proposed hybrid LSTM-GRU with M-AM for text summarization, as the proposed framework is capable of delivering fast and accurate outcome for text summarization. Fig.1 shows the overall method of the proposed work.

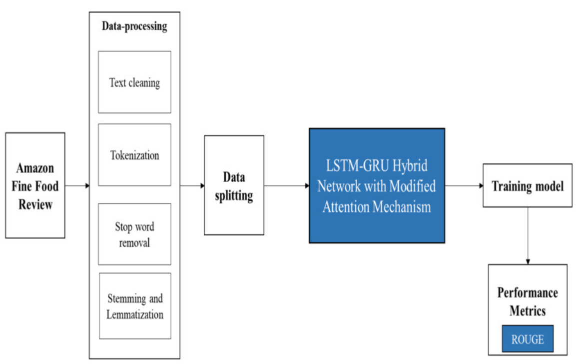

Fig.1: *Overall Method of the Proposed Work.*

The overall flow of the proposed model is depicted in Fig.1. Initially, Amazon fine food review dataset is loaded, then the data present in the dataset is pre-processed using various pre-processing techniques such as Text cleaning, Tokenization, removal of stop words, stemming and lemmatization. Once the data is pre-processed, data is further splitted into train and test splits. Further, the data is fed into the LSTM-GRU hybrid network with a Modified Attention Mechanism (M-AM). LSTM-GRU with the MAM model aids in summarizing a large paragraph into crisp and concise text for an easy and understandable format quickly and accurately. The m-AMm-AM model employed for the attention mechanism uses lexical significance measures and sentence context weights. Finally, the performance of the proposed model is assessed using the ROUGE metric. This metric assists in evaluating the automatic summarization and machine translation software in NLP.

3.1 Data Pre-Processing

Data Pre-processing is employed to remove the redundancies, inconsistencies, and missing values, as these aspects can lead to inaccurate and poor outcomes. Hence, it is extremely important to clean the dataset to avoid the poor user experience. Therefore, the proposed model incorporated various pre-processing techniques such as Text cleaning, tokenization, stop word removal, and stemming & lemmatization.

- Text cleaning aids in the removal of repeated words and other wanted noises, which come in different forms. The objective of implementing text cleaning is to eliminate the dataset's noise while retaining much relevant information.
- Up next, the tokenization process is incorporated. This process aids in breaking the raw text into tiny chunks. Tokenization aids in interpreting the meaning of the text by examining the sequence of the words. Stop word removal aids in removing highly frequent words from the text since it does not add any valuable information to the text. Stop word removal and eliminate common and non-meaningful words such as "the" and "and" from the respective text.
- Finally, stemming and lemmatization are performed for pre-processing, in which stemming aids in reducing the number of unique words and enhances the performance of the model for summarization. Similarly, lemmatization is employed for pre-processing and grouping various inflected forms of similar words.

3.2 Text Summarization – LSTM-GRU Hybrid Network

LSTM is referred to as a type of NN that can be learned with the help of text data. LSTM is measured as a helpful algorithm for a sequence of text since LSTM comprises the running memory, which aids in long-term dependencies and the structure within the long sequences. LSTM with AM (Attention Mechanism) is generally operated by managing the entire output of LSTM within a sequence and training a separate layer to pay attention to specific parts of the output instead of others. The LSTM model is set to return the sequences, in which the sequences of the inputs are $p = (p_1, p_2, \ldots, p_T)$. Moreover, the sequence of the hidden vector is shaped by $q = (q_1, q_2, \ldots, q_T)$. Further, the output is denoted by $h = (h_1, h_2, \ldots, h_T)$of similar length, and the iterating the equations form t=1 to T.

$$h_t = H(Wgt_{ph}p_t + Wgy_{hh}h_{t-1} + b_n) \qquad (1)$$

$$y_t = Wgt_{ht+b_q} \qquad (2)$$

$$h_t = y_t \qquad (3)$$

Here, Wgt is denoted as weight matrices, b is represented as the bias vector, and H is denoted as the function of the hidden layer. Though LSTM can solve the problems of long-term issues such as dependency on RNN. However, the RNN lacked in determining the word stored within the long-term memory. Nevertheless, it could still afford better predictions from the current information. In addition, RNN could deliver suitable performance due to the increase in gap length; however, by default, LSTM has the caliber to sustain the information for a long time. It can predict, process and classify in terms of time series data. Nevertheless, LSTM has a few drawbacks, such as the requirement for more memory due to the usage of additional parameters and operations, which are computationally expensive. Hence, the more advanced model is used as it only possesses fewer gates and fewer parameters than LSTM, which is GRU. Fig.2. shows the architecture of LSTM.

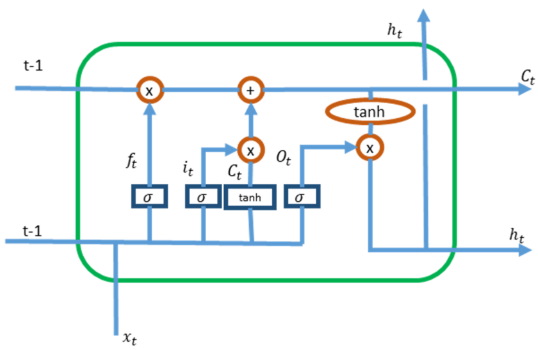

Fig.2: *Architecture of LSTM.*

Where, $t-1$ is denoted as cell state, f_t is represented as Forget gate, h_t is denoted as Hidden gate, i_t represented as input gate, c_t denoted as candidate gate, and o_t is represented as output gate, σ is denoted as sigmoid and tanh is represented as the activation function in Fig.2. The key components of LSTM architecture controls the flow of information, a memory cell to store and update information, and activation functions to control the output.

GRU is considered to be more effective as well as efficient than the LSTM, as GRU possesses the benefits of condensing the structure of LSTM by plunging the computation to update the hidden state through unraveling the concerns of LTD (Long Term Dependency), thereby alleviating the performance of the LSTM. When compared with the LSTM, GRU is faster and employs less consumption of memory as it has the potential to capture long-term dependencies. Due to these reasons, GRU is considered to be more efficient than the LSTM model. Fig.3. shows the architecture of GRU.

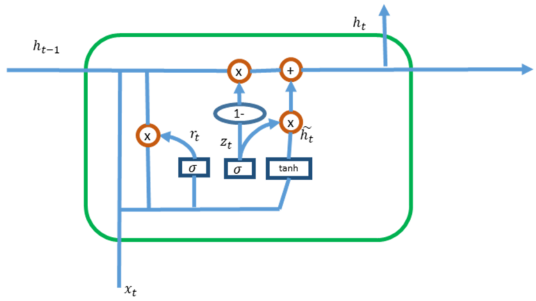

Fig.3: *Architecture of GRU.*

Where z_t is denoted as the update gate, r_t is represented as the reset gate, x_t is denoted as the current time input, H_{t-1} is represented as the previous network state and G_t is denoted as the candidate state in Fig.3. GRU gate is similar to LSTM with a forget gate but it has fewer parameter, however, it consists of fewer parameter than the LSTM model as it does not have an output gate. The performance of GRU is found to be more effective for NLP language. The sequence is initially read by the reset gate, and if the input is considered to be a necessary one, then the input sequence is updated by employing the update gate, which throws the information and waits for the upcoming input. In GRU, the cells are comprised of input and forget gates. In which the input, as well as the forget gate, is controlled by using GC (gate controller) z. input gate is open, and the forget gate is closed if z is denoted as 1, while the input gate is closed, and the forget gate is vice-versa if z is 0.

$$r_t = \sigma(Wgt_r h_{t-1} + U_r p_t) \qquad (4)$$

$$z_t = \sigma(Wgt_z h_{t-1} + U_z p_t) \qquad (5)$$

At every step, previous (t-1) memory is hoarded, and the input of the time step is cleared. Therefore, cells of the GRU are controlled by employing equations 6 and 7,

$$c_t = \tanh(Wgt_c(h_t - 1) + U_c p_t) \qquad (6)$$

$$h_t = (z \otimes c) + ((1 - z)h_{t-1}) \tag{7}$$

GRU is considered an improvement of RNN, which aids as a special gating mechanism to control the overlooking as well as the withholding of information, as it deals with the vanishing of gradients of the variables. Moreover, it also falls into the production of locally ideal solutions. Though GRU has various advantages over LSTM, it is not as effective as LSTM in learning long term dependencies, especially in complicated and complex task. Therefore, LSTM and GRU are combined together and employed for text summarization in the proposed framework.

3.2.1 Modified Attention Mechanism

AM improves the model by selectively focusing on the important input elements by augmenting the accuracy of prediction and aiding in aggregating the computational efficiency. Hence, AM in DL is utilized for concentrating the model on the pertinent part of the input while making the prediction. It challenges to utilize the similar action of selectively concentrating on the pertinent things while snubbing the inappropriate features of DNN. However, AM is considered to be a challenging one for the training process, particularly complicated and large tasks, due to the addition of more weight to the model parameter, which can lead to more increased training time. Due to these reason, it can make the model ineffective and inefficient for text summarization. Fig.4. shows the model of M-AM.

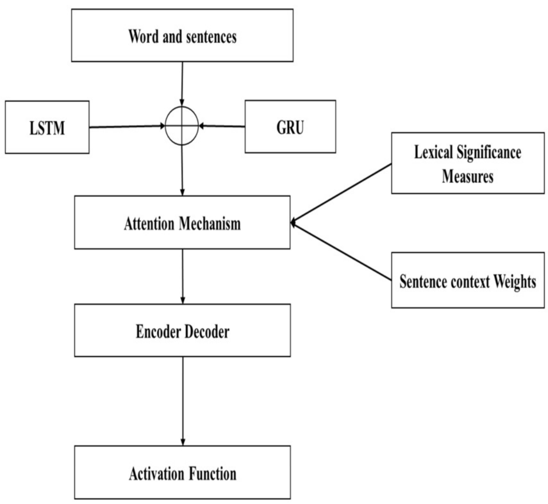

Fig.4: *Modified-Attention Mechanism.*

Therefore, to overcome the problems, the proposed framework employs the M-AM model, which consists of sentence context weights and lexical significance measures. Figure 4 shows the process involved in text summarization. In which the words and sentences are pre-processed, and then the input is fed to the LSTM and GRU model; further attention mechanism is performed using lexical significance measures and sentence context weights. Then, the text enters into the encoder and decoder as the encoder and decoder aid in solving the sequence 2 sequence problem where the input and the output sequences are of various lengths. Lastly, the activation function is utilized as the activation function aids in deciding whether a neuron should be activated or not by calculating the weighted sum and further adding bias to it. The tenacity of the activation function has always been to lead the non-linearity into the output of a neuron.

3.3 Hybrid LSTM-GRU with M-AM

The proposed framework summarizes text to fetch crisp and concise outcomes. Hence, the proposed model employed a hybrid LSTM + GRU model with M-AM for better text summarization. Fig.5 shows the process of the proposed framework for text summarization using LSTM + GRU model.

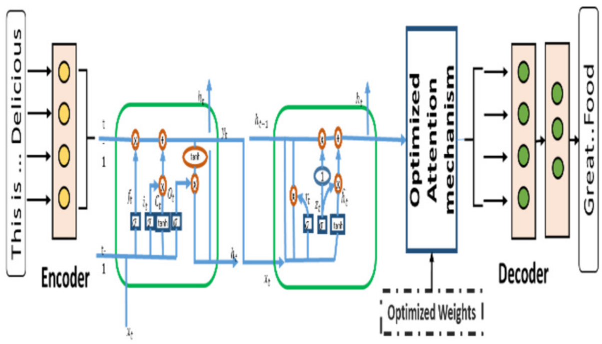

Fig.5: *Mechanism of the Proposed Work.*

Initially, the set of sequences or a sequence is taken as input, and various pre-processing techniques have been taken to remove the unwanted and irrelevant data from the Amazon fine food review dataset. Then, the encoder is used by LSTM as it reads the entire input sequence, wherein, at each step, 1 word is fed into the encoder. Later, it processes the information at every timestep and captures the contextual information that is present in the input sequence. From the encoder, the input is fed to LSTM, in which the input gate decides the relevant information, which could be added from the existing step, and the forget gate determines the relevant and irrelevant data from the previous steps. Finally, the output gate is used for finalizing the next hidden state.

Further, the output from the LSTM is fed to GRU, in which the reset gate is used to read the sequence, and if the input is considered necessary, it is sent to the update gate for updation. Further, an attention mechanism is employed for using M-AM by employing lexical significance measures and sentence context weights for delivering better performance for text summarization. M-AM equations employed in the proposed study,

$$h_t = \tanh(Wgt_p x_t + Wgt_h h_{t-1} + b_h) \tag{8}$$

$$e_t = \sigma(p_t^T Wgt_a x_{t-1} + b_t) \tag{9}$$

$$a_t = \text{softmax}(e_t) \tag{10}$$

In which h_t is denoted as output from the LSTM layer, $_{it}$ is represented as activation output, W_a is represented as the weight of the attention network, $a_{nd\ it}$ is denoted as the softmax activation function. Lastly, the decoder generates the output sequences based on the fixed-length representations attained from the encoder and delivers the output sequence. This working mechanism aids the proposed model in delivering fast and accurate outcomes for text summarization.

4. RESULTS AND DISCUSSION

The proposed design has been executed with Python. The obtained results are discussed in this section, along with a comparative analysis to determine the efficacy of the proposed approach over conventional methods.

4.1 Dataset Description

Amazon's fine food review comprises reviews of fine foods from Amazon. The data span more than ten years, which includes all 5,00,000 reviews till Oct 2012. The review consists of user and product information, and ratings as well as plain text of the review. It encompasses reviews from all categories of Amazon. Table 1 shows the Amazon fine food review dataset.

Table 1: *Amazon Fine Food Review Dataset.*

Data Comprises of
Reviews consist of – 568,454
Users – 74,258
260 Count of users with less than 50 reviews

4.2 Performance Metrics

Performance metrics are employed to evaluate the performance of the projected system by using metrics like ROUGE, accuracy, and loss.

a) ROUGE

ROUGE metric is used to identify employability, which aids in assessing the summaries of the text. It depends on the comparison between reference summaries and generated summaries. It is computed by using the equation 11,

$$\text{ROUGE} = \frac{\sum S \in (\text{reference}_{\text{summaries}}) \sum \text{gram}_n \in S \text{ Countmatchgram}_n}{\sum S \in (\text{reference}_{\text{summaries}}) \sum \text{gram}_n \in (\text{gram}_n)} \tag{11}$$

b) Accuracy

Accuracy is claimed as the measure of total accurate classification. The accuracy range is calculated with the following equation 12,

$$Acc = \frac{TRN + TRP}{TRN + FLN + TRP + FLP} \tag{12}$$

Where TRN signifies True negative, TRP is True positive, FLN is False negative, and FLP is False positive.

c) Loss

The loss is estimated based on the testing and training. The interpretation of the loss is heavily depended on the training and testing performance of the model. Table 2 shows the percentage split of data.

Table 2: *Percentage split of data.*

Set	**Percentage**
Training Set	70% (approximately 350,000 reviews)
Testing Set	15% (approximately 75,000 reviews)
Validation Set	15% (approximately 75,000 reviews)

The data split in Table 2 provides substantial data for training the model while still leaving a considerable portion for testing and validation. Adjustments can be made based on specific requirements, the nature of the task, and the size of the dataset.

4.3 EDA

EDA refers to an approach that analyzes the datasets to summarize the primary characteristics more often with visual approaches. The objective of the EDA is to help look at the data before making any assumptions. The purpose of the EDA is to make any errors and aid in understanding the patterns within the data and identify the outliers and anomalous behavior present in the data. Fig.6 shows the count of the text and the summary count after the data cleaning process, in which the range of text lies between 5000 and 6000, whereas in summary, it lies more than 2000.

4.4 Performance Analysis

Performance of the model is evaluted in the subsequent section with accordance with loss and accuracy of the model. Fig.7 shows the accuracy and loss of the proposed model.

Fig.8 shows that maximum accuracy is obtained in the testing phase, whereas the loss rate, which has

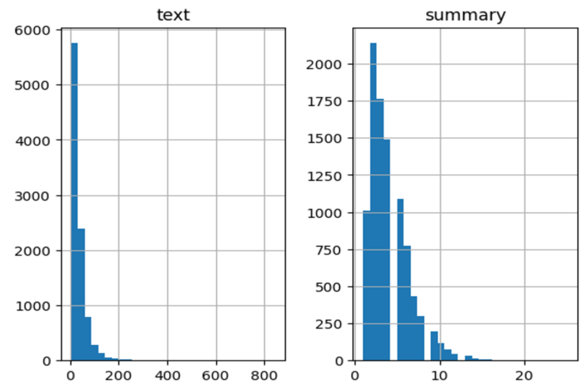

Fig. 6: *Text and summary of the Paragraph.*

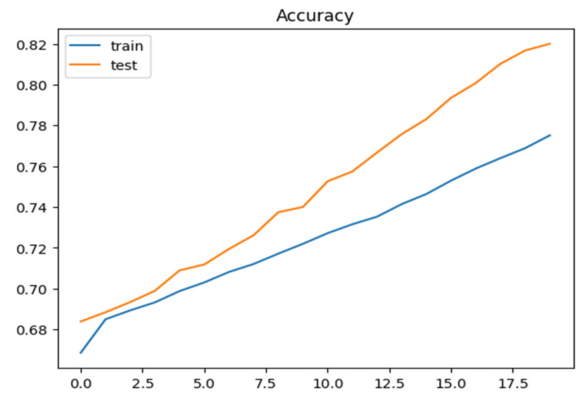

Fig. 7: *Analysis in accordance with accuracy.*

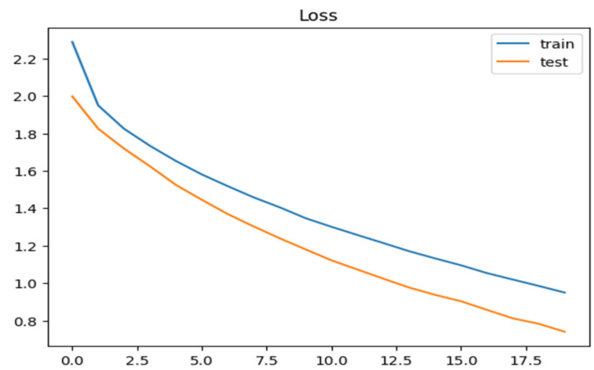

Fig. 8: *Analysis by Loss.*

been attained, was identified to be reduced in the testing stage.

4.5 Experimental Result

Results, which have been obtained after the prediction, are tabulated in Table 3. Evidently, the predicted summary reflected the precise meaning of the original summary without any major changes from the original summary.

Table 3: *Amazon Fine Food Review Dataset.*

Review: fact like artificial taste orange drink taste artificial taste would stick plain orange juice orange like **Original summary:** nothing great about this **Predicted summary:** great product
Review: bought Amazon sold stores delicious taste like cookies bought flavors one like chocolate chip others though wonderful calories try eating healthy treats sometimes perfect treat since never eat regular cookie get tasty cookies **Original summary:** So Delicious **Predicted summary:** Delicious
Review: got tea and decided to try the brand tea. I was surprised by the wonderful flavor, like putting bag, the quality bags value tea. **Original summary:** Great for the price **Predicted summary:** Price good

4.6 Comparative Analysis

Comparative Analysises are used for comparing the conventional methods with the proposed model with the aim of evaluating the efficiency and efficiency of the proposed method. From the table, it can be identified that the existing LSTM has attained the ROUGE rate of 33.33, and the GRU model attains the ROUGE rate of 34.5. Whereas the Proposed model attained a ROUGE rate of 55.5, which shows the efficiency and efficacy of the proposed model for text summarizer. Fig.9 shows the graphical illustration of table 4.

Table 4: *Comparative analysis.*

Performance Metrics	**Rouge-I**
Existing results 1 [29]	33.33
Existing results 2 [30]	34.5
Proposed method	55.5

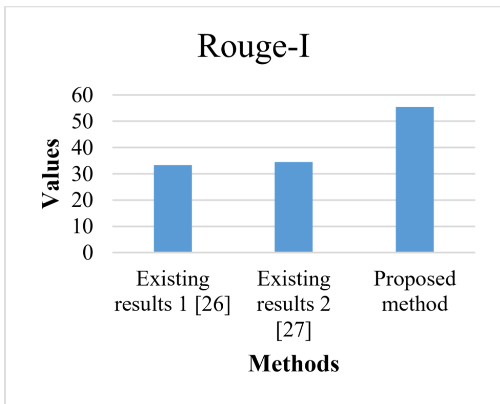

Fig. 9: *Comparative analysis of existing and proposed.*

From the experimental outcome, it was identified that the proposed hybrid LSTM-GRU with the M-AM model had delivered better outcomes than the existing LSTM and GRU methods due to the ability to summarize the text faster and more accurately than the existing ones. The incorporation of M-AM has made the model more efficient for text summarization as these techniques employed sentence context weight and lexical significance measure for attention mechanism. Further, by using M-AM, weights are allotted to the parameters accordingly, making the proposed model efficient and effective for text summarization.

Discussion

The CNN approach is generally considered challenging for capturing long-term context information and needs numerous CNN layers for capturing long-term dependencies. Therefore, the study used a BiLSTM-based CNN model for classifying and reflecting the helpfulness of reviews using the Amazon dataset. However, the model lacks summarizing a text in a specific category [31]. Likewise, the author utilized stacked LSTM based on the Attention mechanism using a sequence-to-sequence model to generate a summary to obtain a short, understandable, and fluent abstractive summary for the Amazon fine food dataset. Nevertheless, the performance of the model was considered to be effective in terms of exactness [32]. Correspondingly, beam search decoder during the inference phase with linear normalization and LSTM in the encoder-decoder sequence–to–sequence model, along with attention mechanism for enhancing the processing speed of the review sentence [33].

Though the studies are capable enough for clear-cut text summarization obtained by the prevailing works, they still lack in delivering satisfactory outcomes due to ineffective algorithms. Therefore, the proposed work predominantly utilized a hybrid LSTM-GRU model with M-AM for fast and precise text summarization. By doing so, the ROUGE metric value obtained by the proposed model is 55.5.

5. CONCLUSION

Text summarization is one of the major domains in NLP. Text summarization is primarily used for creating a short, precise, and crisp text of a longer document. It is considered a beneficial application of NLP, which delivers a short and meaningful summary of a lengthy paragraph, thereby aiding in efficiently understanding the essence of the topic. Therefore, various existing studies have employed different methods for text summarization. However, existing methods do not possess the potential to summarize long text quickly and accurately. To overcome this issue, the proposed model employed Hybrid LSTM-GRU with M-AM for text summarization. The proposed model is pre-processed using text cleaning, tokenization, stop word removal stemming, and lemmatization. This model aided in delivering faster and employs less consumption of memory.

Along with it; it has the potential to capture long-term dependencies as well. Besides, M-AM is employed in the proposed model using lexical significance measure and sentence weight context for an effective text summarization model. Finally, the model's output is identified using the ROUGE metric, in which the value obtained by the proposed model was 55.5%, whereas the ROUGE value obtained by the existing LSTM and GRU was 33.33 and 34.5. In the future, additional, more effective DL algorithms can be used for summarizing the text effectively.

In conclusion, the Amazon Food Review Text Summarization dataset exhibits limitations, including biased representation, potential context oversights, and imbalanced review lengths. Its granularity may be insufficient for nuanced sentiments, and ethical concerns related to user privacy should be acknowledged. The dataset's adaptability to evolving language trends and domain-specific jargon poses challenges. Despite these limitations, leveraging the dataset can yield valuable insights with a mindful approach to address biases and ethical considerations. Further research and refinement are essential for enhancing the dataset's robustness and ensuring its applicability in diverse linguistic and cultural contexts. The current study's limitations include a relatively small sample size, potentially limiting the generalizability of findings. Methodological constraints, such as the chosen research design or data collection methods, may introduce bias or hinder the exploration of certain aspects.

DECLARATION

- **Conflict of Interest:** The author reports no conflict of Interest.
- **Funding:** None
- **Acknowledgment:** None

AUTHOR CONTRIBUTIONS

- **PL. Prabha:** Conceptualization, Methodology, Investigation, Data curation, Formal Analysis, Writing Original Draft
- **Dr.M. Parvathy:** Conceptualization, Methodology, Investigation, Data curation, Formal Analysis, Supervision, Writing & reviewing

References

[1] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. J. E. s. w. a. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Systems with Applications*, vol. 165, no. 113679, 2021.

[2] A. Mahajani, V. Pandya, I. Maria, D. J. A. C. Sharma and C. S. RACCCS-, "A comprehensive survey on extractive and abstractive techniques

for text summarization," *Ambient Communications and Computer Systems*, pp. 339-351, 2019.

[3] S. Shrenika, "Abstractive Multi Document Text Summarization of User Reviews Using Graph Generation and TF-IDF," *Perspectives in Communication, Embedded-systems and Signal-processing-PiCES*, vol. 5, no. 5, pp. 54-57, 2021.

[4] S. Sarica and J. Luo, "Stopwords in technical language processing," *PLoS ONE*, vol. 16, no.8:e0254937, 2020.

[5] D. Khyani, B. Siddhartha, N. Niveditha, and B. Divya, "An interpretation of lemmatization and stemming in natural language processing," *Journal of University of Shanghai for Science and Technology*, vol. 22. no. 10, 2020.

[6] G. Neelima *et al.*, "Extractive text summarization using deep natural language fuzzy processing," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 6S4, pp.2278-3075, 2019.

[7] Z. Fatima *et al.*, "A novel approach for semantic extractive text summarization," *Applied Sciences*, vol. 12, no. 9:4479, 2022.

[8] A. Kovačević and D. Kečo, "Bidirectional LSTM networks for abstractive text summarization," *Advanced Technologies, Systems, and Applications VI*, pp. 281-293, 2021

[9] R. Srivastava, P. Singh, K. Rana, and V. J. K.-B. S. Kumar, "A topic modeled unsupervised approach to single document extractive text summarization," *Knowledge-Based Systems*, vol. 246, no. 108636, 2022.

[10] N.-K. Nguyen, V.-H. Nguyen, and A.-C. Le, "A Local Attention-based Neural Networks for Abstractive Text Summarization," in *Proceedings of the 2023 5th Asia Pacific Information Technology Conference*, pp. 152-159, 2023.

[11] H. Bhathena, J. Chen, and W. Locke, "Recognizing themes in Amazon reviews through Unsupervised Multi-Document Summarization," thisisjeffchen.com. `chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://thisisjeffchen.com/publication/amazon-reviews.pdf`.

[12] A. Pieta, "Use of evolution of deep neural network for text summarization," M.S. thesis, Universitat Politècnica de Catalunya, 2021.

[13] K. Muthiah, "Automatic Coherent and Concise Text Summarization using Natural Language Processing," M.S. thesis, Dublin, National College of Ireland, 2020.

[14] P. M. Hanunggul and S. Suyanto, "The Impact of Local Attention in LSTM for Abstractive Text Summarization," *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, pp. 54-57, 2019.

[15] D. Miller, "Leveraging BERT for extractive text summarization on lectures," arXiv preprint, arXiv:1906.04165, 2019.

[16] R. Atanda, "Text Summarization of Customer Reviews Using Natural Language Processing," Doctoral dissertation, Dublin, National College of Ireland, 2020.

[17] A. K. Mohammad Masum, S. Abujar, M. A. Islam Talukder, A. K. M. S. Azad Rabby and S. A. Hossain, "Abstractive method of text summarization with sequence to sequence RNNs," *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, pp. 1-5, 2019.

[18] D. Debnath, R. Das and S. Rafi, "Sentiment-Based Abstractive Text Summarization Using Attention Oriented LSTM Model," *Intelligent Data Engineering and Analytics*, pp. 199-208, 2022.

[19] K. K. Manore, "Text Summarization of Customer Food Reviews Using Deep Learning Approach," Doctoral dissertation, Dublin, National College of Ireland, 2022.

[20] S. A. Tripathy and A. Sharmila, "Abstractive method-based text summarization using bidirectional long short-term memory and pointer generator mode," *Journal of applied research and technology*, vol. 21, no. 1, pp. 73-86, 2023.

[21] R. Singh, "Text Summarization using Sequence to Sequence," Doctoral dissertation, Dublin, National College of Ireland, 2023.

[22] K. Divya, K. Sneha, B. Sowmya and G. S. Rao, "Text summarization using deep learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 5, pp. 3673-3677, 2020.

[23] Ö. B. Mercan, S. N. Cavsak, A. Deliahmetoglu and S. Tanberk, "Abstractive Text Summarization for Resumes With Cutting Edge NLP Transformers and LSTM," *arXiv preprint*, arXiv:2306.13315, 2023.

[24] N. Sanjabi, "Abstractive text summarization with attention-based mechanism," M.S. Thesis, Universitat Politècnica de Catalunya, 2018.

[25] J. Sheela and B. Janet, "An abstractive summary generation system for customer reviews and news article using deep learning," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 7363-7373, 2021.

[26] F. Boumahdi, A. Zeroual, I. Boulghiti and H. Hentabli, "Generating an Extract Summary from a Document," *Proceedings of the Third International Conference on Information Management and Machine Intelligence*, pp. 103-108, 2022

[27] M. Mohd, R. Jan and M. J. E. S. w. A. Shah, "Text document summarization using word embedding," *Expert Systems with Applications*, vol. 143, no. 112958, 2020.

[28] S. Bhargav, A. Choudhury, S. Kaushik, R. Shukla and V. Dutt, "A comparison study of abstractive and extractive methods for text summarization," *Proceedings of the International Conference on Paradigms of Communication, Computing and Data Sciences*, pp. 601-610, 2022.

[29] K. Pabbi and S. C, "Opinion Summarisation using Bi-Directional Long-Short Term Memory," *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, India, pp. 256-259, 2021.

[30] K. Pabbi, C. Sindhu, I. S. Reddy and B. N. S. J. T. Abhijit, "The Use of Transformer Model in Opinion Summarisation," *Webology*, vol. 18, pp. 1084-1095, 2021.

[31] Q. Li, X. Li, B. Lee and J. J. A. S. Kim, "A hybrid CNN-based review helpfulness filtering model for improving e-commerce recommendation Service," *Applied Sciences*, vol. 11, no. 18:8613, 2021.

[32] M. Singh and V. Yadav, "Abstractive Text Summarization Using Attention-based Stacked LSTM," *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, Sonepat, India, pp. 236-241, 2022.

[33] R. M. Patel and A. J. Goswami, "Abstractive Text Summarization with LSTM using Beam Search Inference Phase Decoder and Attention Mechanism," *2021 International Conference on Communication, Control and Information Sciences (ICCISc)*, Idukki, India, pp. 1-6, 2021.

PL. Prabha is a Research Scholar from the Department of Computer Science and Engineering, Sethu Institute of Technology. I have 14 years of teaching experience. My major area of research include Natural Language Processing, Data Mining and Data Structures.

M. Parvathy received B.E. degree in Computer Science in 2000, M.E. degree in Computer Science in 2009 and Ph.D. degree in Information and Communication Engineering from Anna University, India in 2015. She is working as Professor at Sethu Institute of Technology, Madurai, India. Her major research interest includes data mining, information retrieval, networking and semantic web.