Received: November 13, 2022. Revised: January 2, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

A Sequence-to-Sequence Text Summarization Using Long Short-Term Memory Based Neural Approach

Dipti Yashodhan Sakhare[1]*

[1]*Department of Electronics and Telecommunication, MIT Academy of Engineering, Alandi, India*
*Corresponding author's Email: dysakhare@mitaoe.ac.in

Abstract: The massive development in the different fields of engineering and science brought a significant change in the modern years. In the field of natural language processing, automatic text summarization is known as one of the important research directions. The text's primary concepts and flow should be considered to provide a solid summary that limits the repetition in the text as a summary. A sentence is referred to as a basic language unit and there are many semantic links present among the prison terms like example-of, cause-effect, sequential, etc. within a meaningful text. In this research manuscript, an automated text summarization model is developed and its performance is validated on the two challenging datasets such as daily-mail and Gigaword. In this manuscript, a transformer: long short term memory (LSTM) neural network is used along with the Huber loss function and Adam optimizer. The transformer used in this research (LSTM with Huber loss function and Adam optimizer) includes the advantages like requiring limited memory, computationally effective, and easy implementation. The proposed neural network mainly aims in exhibiting the summaries that are composed by the groups or paragraphs that includes more keywords or phrases than summaries composed by sentences. The non-differential evaluation metrics are utilized in the proposed sequence-to-sequence model to provide semantic information of input text and it stores important features for effective text summarization. The size of the parameters is specified by the maximum number of sentences in the same group. The proposed LSTM with Huber loss function and Adam optimizer has 44.51 ROUGE-1, 20.43 ROUGE-2, and 40.08 ROUGE-L on dailymail dataset. Experimental results of the trial demonstrated that the proposed model outperformed the existing models like dynamic residual network, convolutional neural network with LSTM, and knowledge powered topic level attention model in terms of the rouge parameter.

Keywords: Natural language processing, Long short-term memory neural network, Semantic representation techniques, Sequence to sequence models, Text summarization.

1. Introduction

Automatic text summarization systems are increasingly in demand as a result of the web's rapid expansion in textual data and the difficulty in locating desired information among the massive volume of data. [1]. Natural language processing plays a significant role in automatic text summarization and the extracted sentences are usually coherent [2]. An important area of research in natural language recognition is machine translation synthesis and considered the best approach to specify the growth of online texts. Texts should convey a good summary fluently while minimizing the redundancy of the deep ideas of the source [3]. Obtaining the required information from a huge amount of information is one of the difficult tasks for scientists and for this, the taxonomy of text summarization is divided into multi or single document summarization tasks [4]. In this proposed method, sequence to sequence models is implemented for text summarization [5-7].

In recent times, the development of automatic summarization is increased worldwide. The document type or domain features aren't considered in most of the text summarization approaches [8]. Hence, this text summarization works with the direct components acquired from the document such as sentences, paragraphs, terms, and so on [9]. Because of their promising performance, neural networks are now frequently used in various natural language processing applications. In text summarization jobs, the encoder reads the whole input sequence and creates a fixed-dimensional feature vector, which is then used by the decoder to build the required output sequence. In the summarization model, a CNN serves as the encoder and a language model using a collection of computational models serves as the decoder. The complexity of the natural language processing (NLP) area, as well as our overall lack of knowledge of natural languages and human intellect, are the key obstacles to producing such an assessment. In this study, the research focus on one particularly relevant and difficult aspect of NLP evaluation: how to evaluate the content of two text sections semantically [10]. The objectives and contribution of this research are given as follows:

- The sequence-to-sequence model's summarization performance is estimated based on non-differentiable evaluation metrics. The proposed sequence-to-sequence model includes two phases like encoder and decoder that are used for text summarization.
- The proposed LSTM neural network with Huber loss function and Adam optimizer selects optimal features from input data and performs summarization. The proposed LSTM model has long-term dependency that can store relevant information for the long term.
- The proposed LSTM neural network with Huber loss function and Adam optimizer performance is validated on Daily Mail and Gigaword datasets. The proposed model obtains higher performance in text summarization than the existing techniques. The use of various clustering algorithm has an impact on group creation. A significant contribution is made by this work in the extraction of text summarization and in verifying the role of semantic link networks and in representing and understanding the texts.

The paper is organized as follows: the analysis of available methods based on the sequence-to-sequence ranking is reviewed in section 2. In Section 3, the suggested model is mathematically described. Section 4 describes the experimental findings and analysis, and section 5 details conclusion of the proposed model.

2. Literature review

Rouane [11] have introduced a novel biological data summary approach that incorporates grouping and frequent pattern processing. Biomedical papers were used as a set of biomedical concepts using the UMLS met thesaurus. The resulting summaries with the abstract of text documents use the ROUGE metrics in terms of recall, precision, and F-measure. To decrease delicate information and to upgrade the standard of summaries, an anti-redundancy technique should be incorporated as a future extension. Mojrian and Mirroshandel [12] introduced a multi-document text summarization approach known as MTSQIGA. In this literature study, a modified quantum-inspired genetic algorithm (QIGA) was used to handle some stages to discover the top answers. MTSQIGA was compared with existing state-of-the-art and specifies the hopeful effectiveness of the proposed algorithm. However, the QIGA was extremely dependent on the optimal of its hyper-parameters value and individual performance, where it was a computationally impulsive task.

Cao and Zhuge [13] implemented an extractive text summarization, which mainly aims at extracting the important sentences, where it was related to many groups. The results of the experiments shows that the summaries composed by groups or paragraphs contain many numbers of phrases or keywords than the summaries composed by sentences; the maximum number of sentences that were present in a group was specified by the size of the parameters. The comparison of the seven clustering algorithms was used to generate collections made by CNN-DM and Gigaword datasets. Still, the developed model has the problems in identifying the text, interpretation, and evaluation of the generated summary. Lierde and Chow [14] implemented a creative methodology for identifying the document subjects and recovering sentences from documents that were pertinent to a given query. It draws on the influential notion of graph based transversals. The current study uses graph-based ranking algorithm to calculate each sentence's unique score, but it fails to account for the themes that were jointly covered by the phrases, while creating a summary, where it leads to repeated summaries.

Mohamed and Oussalah [15] has implemented a novel framework for both single and multi-document summarization and it was built based on graph. The summarization system was evaluated using CNN-DM and Gigaword datasets. The outcome shows significant performance gains in summary quality, demonstrating the potency of the position representation of knowledge and its linking. This research study does not possess the sequence of sentences obtained from the semantic representation of the summary, which was considered as a major issue. Aliakbarpour [16] applied an abstractive

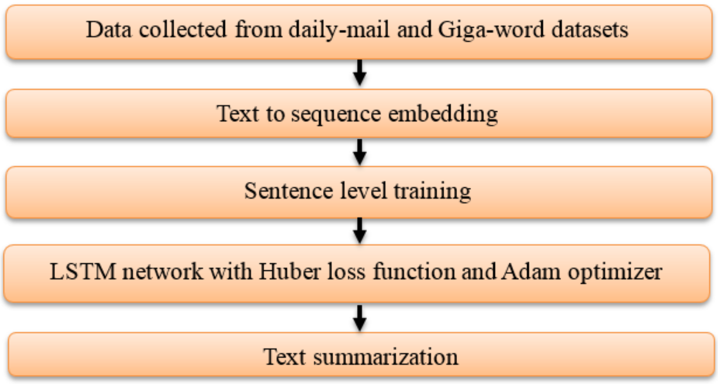

Figure. 1 Block representation of text summarization

summarization technique for auxiliary attention (AA) with convolutional neural network (CNN) and LSTM combination to increase saliency and generated summaries coherency. The AA-CNN-LSTM model was applied to evaluate the CNN/Daily mail dataset and tested in terms of ROUGE metrics. The combination of CNN and LSTM was applied in the encoder to use phrases as input and generate natural sentences as summaries. Auxiliary attention was used in the encoder and outlining with filtering the information was used for simulating the summary generated, but the process was computationally complex and costly.

Liao [17] applied an encoder-decoder model according to a variable recurrent neural network for the limitation of long-term dependence. The model dynamically selects the optimal state from state history for connection establishment to increase the dependencies of a long sequence of LSTM related to the current decoding environment. To simulate word reliance, dynamic remnant connectivity and supervised learning were used for long-term connection-dependent words. Hence, the possibility of missing relevant information was higher in this study, which needs to be addressed as a future work. Khanam [18] applied knowledge powered topic level attention (KTOPAS) model based on a convolutional sequence network of text summarization model and topic knowledge base (TKB). The topic model was used to provide coherent and insightful topic information based on knowledge, and the TKB method was used to retrieve conceptual pertinent wisdom. Topic information of knowledge power in KTOPAS from TKB and topic knowledge was applied in a convolutional sequence network. However, the KTOPAS model was highly dependent on context, which was inappropriate in the real time practical applications. In order to highlight the above stated concerns, a new transformer: LSTM with Huber loss function and Adam optimizer is proposed in this manuscript for effective text summarization.

3. Proposed methodology

Sequence to sequence ranking is the establishment of academics that provides creative ideas in its implementation. In this proposed work, the content knowledge of classification is done by sequence-to-sequence ranking model: LSTM with Huber loss function and Adam optimizer. The block diagram of the proposed work is shown in Fig. 1.

3.1 Dataset collection

The CNN-DM and Gigaword are two challenging datasets used in this article. HH-ATS is measured in the CNN-DM dataset, which is used to evaluate the models such as ATS. CNN-DM dataset has broadcasting objects and some summaries are written by humans [19]. The 287226 training samples are collected in total. On average, 781 tokens are present in each article and 56 tokens are present in each summary. Gigaword is a dataset that is spontaneously formed by making use of the initial verdict of the respective article [20]. Around 3.8M training sets, 400 thousand testing sets, and also 400 thousand validation sets are present in total in the Gigaword dataset as text articles. A sample article of the dataset is shown in Fig. 2.

3.2. Preprocessing

After collecting the datasets, the process of data pre-processing takes place using statistical or linguistic techniques. Linguistic and Statistical techniques are the two extractive methods that usually rank the sentences by analyzing parts of speech of phrases or words, word or phrase frequency, the position of a sentence, etc.

Eq. (1) shows the semantic rank presentation, the semantic keywords, and document collection belonging to the preprocessing step, all n-grams of

HAMILTON, Bermuda (CNN) -- Four Chinese nationals of Uyghur ethnicity who had been held at the U.S. military's Guantanamo Bay, Cuba, detention facility have been resettled in Bermuda, officials said Thursday.
Attorney General Eric Holder says the U.S. is "extremely grateful to the government of Bermuda."
"Above all, this was a humanitarian act," Bermudan Premier Ewart Brown told CNN in an interview at his Cabinet office in Hamilton, Bermuda. "We don't see it as quid pro quo."
The four were twice cleared for release -- once by the Bush administration and again this year, according to a Justice Department statement.
They were among 17 Uyghur detainees at the facility set up to hold terror suspects.
The four were flown by private plane Wednesday night from Cuba to Bermuda and were accompanied by U.S. and Bermudan representatives as well as their attorneys, according to Susan Baker Manning, part of the men's legal team.
President Obama has pledged to close the Guantanamo facility, raising questions of what will happen to the more than 200 remaining detainees. A political backlash against bringing any of the detainees to the United States has increased the focus on sending them to other countries.
Brown said he read an article on the issue of the Guantanamo Bay detainees' fates in The Washington Post while he was in Washington for a White House meeting in May. He said he decided to put an offer to the U.S. government "on the table."
He said Bermuda, a British colony, told London of its intentions, but not until late in the process. Britain must approve the transfer for it to be permanent, Brown said, adding that he believes the issue may raise tension between Bermuda and Britain.
The issue is controversial because of China's opposition to the Uyghurs being sent to any country but China.

Figure. 2 Sample text article

size up to five words using a dictionary lookup (WordNet and Wikipedia) and sliding windows are detected by the algorithms. The set of distinct sentences is represented in Eq. (1). In case of ω_{ij} the formula is given by,

$$\omega_{ij=}\lambda_{ti\,tj}\,.SRT(t_{i\,,}t_j) \tag{1}$$

Where, ω_{ij} is the edge weights, SRT is the Semantic Rank Uses, and t_i and t_j are the terms in between semantic relatedness, which is given in Eq. (2)

$$SRT\,(t_{i\,,}t_j) = \begin{cases} 1, & t_i = t_j \\ SR_{WN(t_{i\,,}t_j)}, & if\ t_{i\,,}t_j\ \in Word\ net \\ SR_{WiKi(t_{i\,,}t_j)}, & if\ t_{i,}t_j\ \in WiKipedia \\ 0, & Otherwise \end{cases} \tag{2}$$

Two measures are combined in a single measure SRT (t_i, t_j), as shown in Eq. (2)

The schematic relatedness between the two terms t_i and t_j according to WLM is defined as shown in Eq. (3).

$$SR_{WiKi}(t_i\,, t_j) = \frac{log(max\{|In(a_i)|,|In(a_j)|\})-log(|In(a_i)\cap In(a_j)|)}{log(|w|)-log(min\,\{|In(a_i)|,|a_j|\})} \tag{3}$$

The measure presented between a pair of terms is shown in Eq. (4). The semantic relatedness between two terms t_i andt_j, when $t_i \in w_N$ and $t_j \notin w_N$ does not belong to or vice versa is considered as 0. It is represented in Eq. (4).

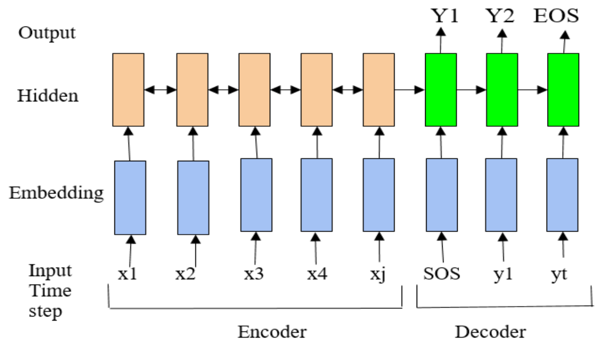

Figure. 3 The basic sequence 2 sequence model

$$SR_{W_N}(t_i, t_j) = max_m\{max_k\{SCM(S^m_{ij}, P^k_{ij}).SPE(S^m_{ij}, P^k_{ij})\}\} \tag{4}$$

SCM - Semantic compactness
SPE – Semantic path elaboration.

3.3 Feature extraction

After preprocessing the data, the process of feature extraction is carried out using assumptions that are based on summaries. Two classes of models are developed based on both word and sentence retrieval. The model which is developed can be informed by informativeness attributes and trained on massive datasets. Neural abstractive text summarizer (NATS) is used which is a method of creating shorter, semantically relevant phrases that convey the substance of the source text's overall meaning, is employed.

3.4. Sequence to sequence model

In this proposed method, sequence-to-sequence models is implemented for text summarization. Non-differentiable evaluation metrics estimate the performance of sequence-to-sequence summarization. The basic sequence 2 sequence model is depicted in Fig. 3.

Sequence 2 sequence framework text summarization is largely collected of an encoder and a decoder. Articles are read by the encoder and denoted by $x = (x1, x2, \ldots, xj)$ and transformed into hidden states.

$y = (y1, y2, \ldots, yt)$. x_i , y_j - symbols of the tokens in the source article.

The definition of a synthesis work is deducing a summary y derived from an original report x and it is shown in Fig. 1 and the initialized equation is shown in Eq. (5).

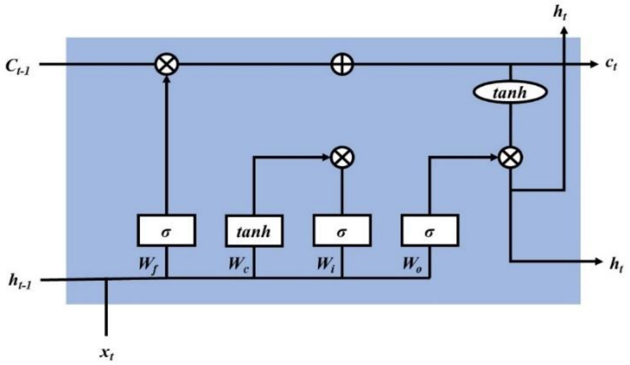

Figure. 4 The LSTM unit cell

$$h_o^d = \tan h(W_{e2d}(\overrightarrow{h_J^e} \oplus \overleftarrow{h_1^e}) + b_{e2d}), C_o^d = \overrightarrow{C_J^e} \oplus \overleftarrow{C_1^e} \quad (5)$$

A decoder is denoted by 'd' and a concatenation operator is represented by the symbol $\oplus$.

3.5 LSTM neural network

To execute the gradient problems, Long short term memory (LSTM) [21-22] is the best solution and exploding gradients issues are not addressed. The LSTM unit cell diagram is shown in Fig. 4. Some mechanisms for abstractive text summarization are discussed as follows:

The actual report x to the target yt is dependent on the attentiveness quotient α_t^e are given by location softmax. The switching network estimates the $P_{gen,t}$ probability of producing the tokens from Z_t^e, a vocabulary content and h_t^d, a hidden state which is represented in the equation in Eq. (6).

$$P_{gen,t} = \sigma(W_{s,z}Z_t^e + W_{s,h}h_t^d + b_s \quad (6)$$

$P_{gen,t}$ - Scalar quantity, $\sigma(a) = \frac{1}{1+\exp(-b)}$ - sigmoid activation function.

Switching generator-pointer - Generation of the symbol from terminology or a point to the source at a specific translating stage is determined as a switching generator that is equipped by the switch. The switch is expressed in Eq. (7),

$$P_{gen,t} = \sigma(W_{s,z}Z_t^e + W_{s,h}h_t^d + W_{s,e}E_{yt-1} + b_s \quad (7)$$

In case when the switch is turned on, a word is produced by a decoder from the vocabulary with the distribution $P_{vocab,t}$ or else α_t^e is generated by the decoder.

i.e. $P_j = \arg max_{j\epsilon\{1,2,\ldots,J\}} \alpha_{tj}^e$.

P_j - Placement of the emblem within the original report.

Copy Net – Different network architectures are present in a copy net end-to-end manner. Copy net tokens are represented by ν and χ, and $V_{ext} = \nu \cup \chi \cup$ <unk> is the extended vocabulary. The vocabulary distribution is represented in Eq. (8). The hyper-parameters used in the LSTM network are loss function: Huber loss, optimizer: Adam, units: 64, and LSTM layers: four.

$$P_{v_{ext}}(yt) = Pg(y_t) + P_c(y_t) \quad (8)$$

3.6 Summarization

Sequence to sequence ranking is applied in an automated summary extraction task. The task for the participating systems in all the competitions is to propose a summary of at most 100 words for each document. The ROGUE toolkit is used for the evaluation against the reference summaries. CNN-DM and Gigaword both datasets comprise a huge number of new articles. For both datasets, two reference summaries per document are provided in the case of both datasets. Therefore, in the cases of both datasets, the proposed system is present among the top two systems in the task.

4. Experimental result and discussion

In this section, the experimental results of the proposed method are described by sequence-to-sequence ranking using LSTM. The validation of the proposed method is carried out with the collected datasets. The proposed text summarization method is applied on a computer with 2.2 GHz and 8GB RAM. The performance matrix and the performance analysis for this method are against the existing approaches which are explained as follows:

4.1 Performance metrics

By using sequence to sequence ranking with LSTM based encoder-decoder framework in which the proposed work considered 70% of data taken for training and 30% of the data taken for testing and the proposed method. ROUGE metrics contain ROUGE-N (N $\in \{1,2,3,4\}$) and ROUGE-L. ROUGE-1, ROUGE-2, ROUGE-3, ROUGE-4 and ROUGE-L are the scores connected to form a R vector. $R_S = R1_S, R2_S, R3_S, R4_S, RL_S$ where R_S is the ROGUE vector. ROGUE is compared with existing methods and the designed four metrics are known as L2 metrics, Increase metric, Increase % metric and divergence metric which are shown in Eqs. (9–12) respectively.

Received: November 13, 2022. Revised: January 2, 2023.

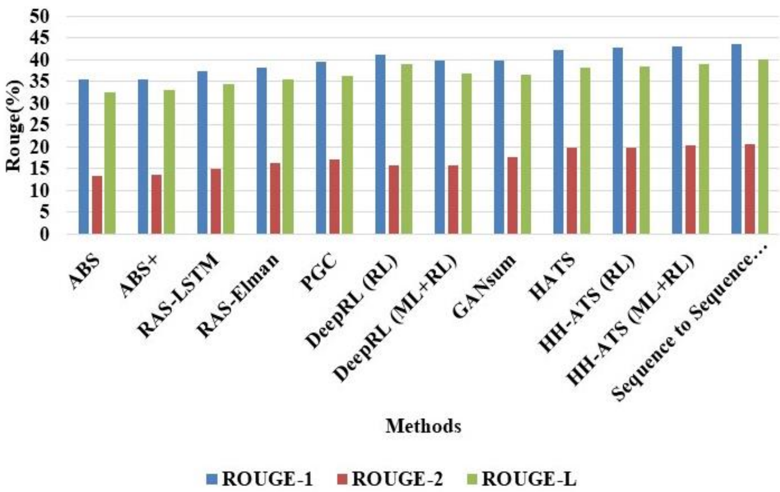

Figure. 5 Comparative analysis of daily mail dataset

Table 1. Automatic evaluation daily mail dataset

Strategies	ROUGE-1	ROUGE-2	ROUGE-L
RAS-LSTM	37.46	15.11	
RAS-Elman	38.25	16.28	34.45
PGC	39.53	17.28	35.43
DeepRL (RL)	41.16	15.82	36.38
ABS	35.46	13.30	39.08
ABS+	35.63	13.75	32.65
DeepRL	39.87	15.82	33.01
(ML+RL)	39.92	17.65	36.90
GANsum	40.74	18.14	36.71
MATS	42.16	19.85	37.15
HATS	42.93	19.85	38.35
HH-ATS (RL)	43.16	20.32	38.64
HH-ATS			39.14
(ML+RL)			
Sequence to Sequence ranking using LSTM	43.52	20.71	40.15

- **L2 metric**

$$\text{L2 } (\text{R}\|R_S) = \sqrt[2]{\sum_{j\in\{1,2,3,4,L\}}(Ri)^2} - \sqrt[2]{\sum_{j\in\{1,2,3,4,L\}}(Ri)^2} \quad (9)$$

- **(IM)- Increase Metric**

$$\text{Increase } (\text{R}\|R_S) = \sum_{j\in\{1,2,3,4,L\}}(Ri - Ri_S) \quad (10)$$

- **IM %** – Usually ROGUE-1, and ROGUE-L scores are somewhat more developed than ROGUE-3 and ROGUE-4 scores. In the statistic for development and L2 metric, a substantial improvement in R3 and R4 can be easily observed and also a small decrease in R1 and RL.

$$\text{Increase \% } (\text{R}\|R_S) = \frac{\sum_{j\in\{1,2,3,4,L\}}\frac{Ri-Ri_S}{Ri_S}*100\%}{5} \quad (11)$$

- **Diverge metric** – Diverge (R‖ R_S) is used to measure the ROUGE vector R_S which is considered a basic ROUGE vector as shown in Eq. (12).

$$\text{Diverge } (\text{R}\|R_S) = \sum_{j\in\{1,2,3,4,L\}} Ri * In\frac{Ri}{Ri_S} \quad (12)$$

Where, Ri_S - F-score of ROUGE-j, j ({j ∈ $1,2,3,4,L$})
Rp, Rg - ROUGE vectors.
Rs - a reference point for contrast
R – ROUGE an outline of the synopsis.

Fig. 5 and Table 1 describe the results of existing and proposed methods for the text summarization evaluation of the Daily mail dataset with Automatic evaluation. Table 1 shows that the proposed Sequence to Sequence LSTM model has better results than the other methods. Sequence to Sequence LSTM model has Rouge-1, Rouge-2 and Rouge-L as 43.52, 20.71 and 40.51 as summarization measures respectively which is better than the existing methods like Deep Reinforcement Learning (DeepRL and DeepML+RL) and MATS method.

Table 2 and Fig. 6 describe the results of existing and proposed methods for the text summarization evaluation of Gigaword dataset with automatic evaluation. Table 2 shows that the proposed sequence to sequence LSTM model has better results than the other methods. Sequence to sequence LSTM model has Rouge-1, Rouge-2 and Rouge-L as 40.23, 20.12 and 36.75 as summarization measures respectively which are better than the existing methods like deep

 DOI: 10.22266/ijies2023.0430.12

Received: November 13, 2022. Revised: January 2, 2023.

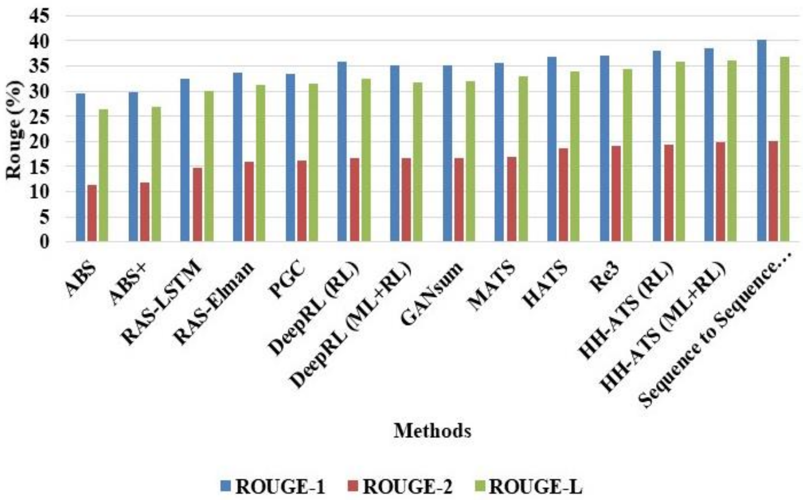

Figure. 6 Evaluation of Gigaword dataset

Table 2. Automatic evaluation of Gigaword dataset

Strategies	ROUGE-1	ROUGE-2	ROUGE-L
ABS	29.55	11.32	26.42
ABS+	29.78	11.89	26.97
RAS-LSTM	32.55	14.70	30.03
RAS-Elman	33.78	15.97	31.15
PGC	33.44	16.09	31.43
DeepRL (RL)	35.82	16.64	32.45
DeepRL	35.16	16.75	31.68
(ML+RL)	35.04	16.55	31.96
GANsum	35.56	16.97	32.94
MATS	36.78	18.65	33.96
HATS	37.04	19.03	34.46
Re^3	38.06	19.28	35.82
HH-ATS (RL)	38.43	19.75	36.11
HH-ATS			
(ML+RL)			
Sequence to Sequence ranking using LSTM	40.23	20.12	36.75

reinforcement learning (DeepRL and DeepML+RL) and MATS method.

Table 3 and Table 4 describes existing and proposed methods results for the text summarization evaluation of the daily mail and giga word dataset with ablation evaluation. Table 3 shows that the proposed sequence to sequence LSTM model has better results than the other methods. Sequence to sequence LSTM model has Rouge-1, Rouge-2 and Rouge-L as 44.51, 20.43 and 40.08 as summarization measures respectively which is better than the existing methods like Deep Reinforcement Learning (DeepRL and DeepML+RL) and MATS method.

The sequence to sequence ranking using LSTM model with Huber loss function and adam optimizer

Table 3. Ablation evaluation of regular mail data source

Strategies	ROUGE-1	ROUGE-2	ROUGE-L
HH-ATS	43.16	20.32	39.14
W/O KB	42.52	19.56	38.25
W/O text	41.74	19.08	38.21
W/O syntax	42.83	19.82	38.60
W/O GAN	41.69	19.05	37.86
W/O syntax	41.32	18.64	37.97
+text	40.43	18.10	37.15
W/O KB +	40.61	18.26	37.31
syntax+ text	40.15	17.90	36.84
W/O GAN + KB			
W/O GAN+text			
+syntax			
Sequence to Sequence ranking using LSTM	44.51	20.43	40.08

Table 4. Ablation evaluation of Gigaword dataset

Strategies	ROUGE-1	ROUGE-2	ROUGE-L
HH-ATS	38.43	19.75	36.11
W/O KB	37.73	19.28	35.36
W/O text	37.85	19.05	35.42
W/O syntax	38.21	19.37	35.80
W/O GAN	37.52	18.59	34.75
W/O syntax+	37.45	18.71	35.07
text	36.70	17.83	33.85
W/O KB + text	36.12	17.63	33.63
+ syntax	35.75	17.37	33.24
W/O GAN + KB			
W/O GAN +			
text + syntax			
Sequence to Sequence ranking using LSTM	38.91	20.07	36.45

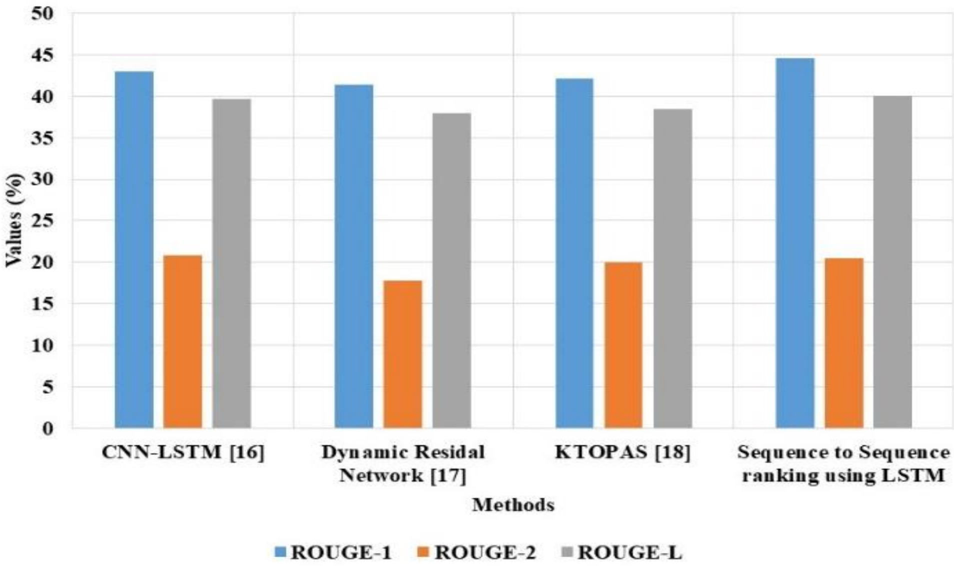

Figure. 7 Comparative analysis of DailyMail dataset

Table 5. Comparison analysis on DailyMail dataset

Model	ROUGE-1	ROUGE-2	ROUGE-L
CNN-LSTM [16]	42.93	20.78	39.63
Dynamic Residal Network [17]	41.35	17.73	37.91
KTOPAS [18]	42.1	20.01	38.45
Sequence to Sequence ranking using LSTM	44.51	20.43	40.08

is evaluated on Daily Mail dataset and compared with existing techniques, as shown in Table 5 and Fig. 7. The sequence-to-sequence ranking model with LSTM model with Huber loss function and Adam optimizer uses the non-differential evaluation metric to encode the data. This helps to provide the semantic information in the sequence-to-sequence model and LSTM model stores relevant features for summarization. The sequence-to-sequence ranking using LSTM model with Huber loss function and Adam optimizer achieved 44.51 ROUGE-1, 20.43 ROUGE-2, and 40.08 ROUGE-L, where the obtained results are better compared to the existing models: CNN-LSTM [16] Dynamic Residal Network [17], and KTOPAS [18].

The sequence-to-sequence ranking using LSTM model with Huber loss function and Adam optimizer is evaluated on Gigaword dataset and compared with the existing KTOPAS [18], as mentioned in Table 6 and Fig. 8. The non-differential evaluation metric is used to encode the data in sequence to sequence ranking LSTM model. This technique helps to provide the semantic information of input data and increases the performance of LSTM in summarization. The sequence-to-sequence ranking with the LSTM model has 38.91 ROUGE-1, 20.07 ROUGE-2, and 36.45 ROUGE-3, the KTOPAS [18] has 37.85 ROUGE-1, 18.71 ROUGE-2, and 33.96 ROUGE-L on Gigaword dataset. The experimental results showed that the proposed model achieved better performance in text summarization and addressed the issues mentioned in the literature section like proper identification of the text, interpretation, and evaluation of the generated summary.

5. Conclusion

The summaries composed of sentences are compared to the synopsis made up of larger layers that have a similar ROUGE score. The group of sentences are combined along original message’s length produces the quality summary in this research. The results demonstrate that the suggested model surpasses cutting-edge baseline techniques in terms of ROUGE ratings. The sentences-to-sentences LSTM model may produce highlights with improved information and elegance, according to the human evaluation as well. The no differential evaluation metric is applied in sequence and sequence based models to provide semantic information of input data. The sentences to sentences using the LSTM model has 38.91 in ROUGE-1, 20.07 in ROUGE-2, and 36.45 in ROUGE-L on Gigaword dataset, existing KTOPAS model has 37.85 in ROUGE-1, 18.71 in ROUGE-2, and 33.96 in ROUGE-L. Future work can propose many patterns of linguistic text summarization which usually ranks the sentences by analyzing parts of speech of phrases or words.

Table 6. Comparison analysis on the Gigaword dataset

Model	ROUGE-1	ROUGE-2	ROUGE-L
KTOPAS [18]	37.85	18.71	33.96
Sequence to Sequence ranking using LSTM	38.91	20.07	36.45

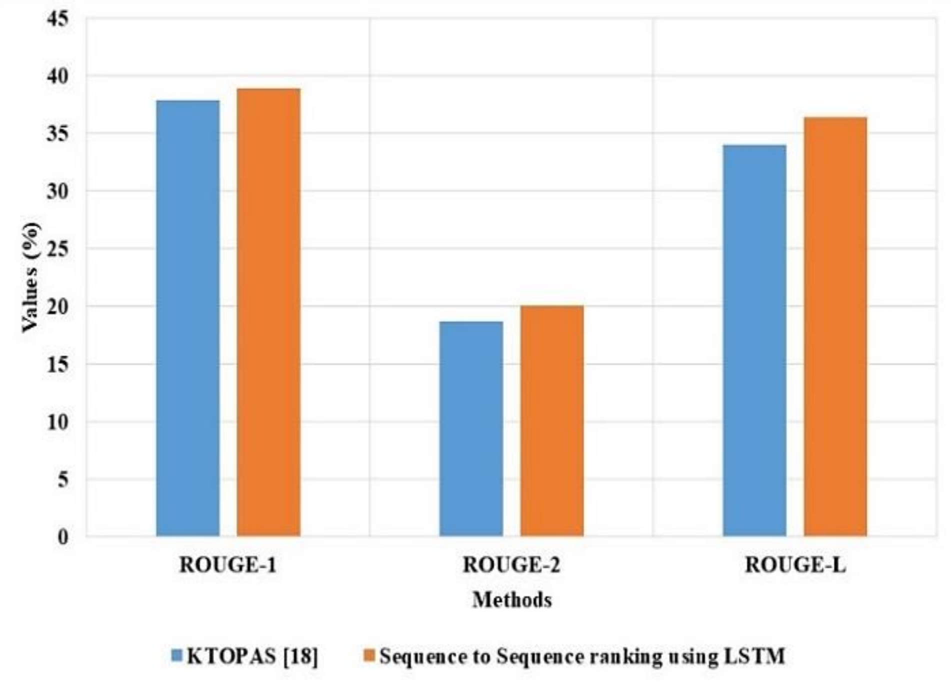

Figure. 8 Comparative analysis of Gigaword dataset

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

The paper background work, conceptualization, methodology, Dataset collection, implementation, result analysis and comparison, preparing and editing draft, visualization, the supervision, review of work and project administration, have been done by first author.

References

[1] R. Sahba, N. Ebadi, M. Jamshidi, and P. Rad, "Automatic text summarization using customizable fuzzy features and attention on the context and vocabulary", *World Automation Congress (WAC)*, pp. 1–5, 2018.

[2] A. Joshi, E. Fidalgo, E. Alegre, and L. F. Robles, "SummCoder: An unsupervised framework for extractive text summarization based on deep auto-encoders", *Expert Systems with Applications*, Vol. 129, pp. 200–215, 2019.

[3] A. Gidiotis and G. Tsoumakas, "A divide-and-conquer approach to the summarization of long documents", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 28, pp. 3029–3040, 2020.

[4] F. B. Goularte, S. M. Nassar, R. Fileto, and H. Saggion, "A text summarization method based on fuzzy rules and applicable to automated assessment", *Expert Systems with Applications*, Vol. 115, pp. 264–275, 2019.

[5] Q. A. A. Radaideh and D. Q. Bataineh, "A hybrid approach for arabic text summarization using domain knowledge and genetic algorithms", *Cognitive Computation*, Vol. 10, No. 4, pp. 651–669, 2018.

[6] C. F. Tsai, K. Chen, Y. H. Hu, and W. K. Chen, "Improving text summarization of online hotel reviews with review helpfulness and sentiment", *Tourism Management*, Vol. 80, p. 104122, 2020.

[7] T. Uçkan and A. Karcı, "Extractive multi-document text summarization based on graph independent sets", *Egyptian Informatics Journal*, Vol. 21, No. 3, pp. 145–157, 2020.

[8] T. Shi, Y. Keneshloo, N. Ramakrishnan, and C. K. Reddy, "Neural abstractive text summarization with sequence-to-sequence models", *ACM Transactions on Data Science*, Vol. 2, No. 1, pp. 1–37, 2020.

[9] R. M. Alguliyev, R. M. Aliguliyev, N. R. Isazade, A. Abdi, and N. Idris, "COSUM: Text

summarization based on clustering and optimization", *Expert Systems*, Vol. 36, No. 1, p. e12340, 2019.

[10] K. Yao, L. Zhang, D. Du, T. Luo, L. Tao, and Y. Wu, "Dual encoding for abstractive text summarization", *IEEE Transactions on Cybernetics*, Vol. 50, No. 3, pp. 985–996, 2020.

[11] O. Rouane, H. Belhadef, and M. Bouakkaz, "Combine clustering and frequent item sets mining to enhance biomedical text summarization", *Expert Systems with Applications*, Vol. 135, pp. 362–373, 2019.

[12] M. Mojrian and S. A. Mirroshandel, "A novel extractive multi-document text summarization system using quantum-inspired genetic algorithm: MTSQIGA", *Expert Systems with Applications*, Vol. 171, p. 114555, 2021.

[13] M. Cao and H. Zhuge, "Grouping sentences as a better language unit for extractive text summarization", *Future Generation Computer Systems*, Vol. 109, pp. 331–359, 2020.

[14] H. V. Lierde and T. W. S. Chow, "Query-oriented text summarization based on hypergraph transversals", *Information Processing & Management*, Vol. 56, No. 4, pp. 1317–1338, Jul. 2019.

[15] M. Mohamed and M. Oussalah, "SRL-ESA-TextSum: A text summarization approach based on semantic role labeling and explicit semantic analysis", *Information Processing & Management*, Vol. 56, No. 4, pp. 1356–1372, Jul. 2019.

[16] H. Aliakbarpour, M. T. Manzuri, and A. M. Rahmani, "Improving the readability and saliency of abstractive text summarization using combination of deep neural networks equipped with auxiliary attention mechanism", *The Journal of Supercomputing*, Vol. 78, No. 2, pp. 2528-2555, 2022.

[17] W. Liao, Y. Ma, Y. Yin, G. Ye, and D. Zuo, "Improving abstractive summarization based on dynamic residual network with reinforce dependency", *Neurocomputing*, Vol. 448, pp. 228-237, 2021.

[18] S. A. Khanam, F. Liu, and Y. P. P. Chen, "Joint knowledge-powered topic level attention for a convolutional text summarization model", *Knowledge-Based Systems*, Vol. 228, p. 107273, 2021.

[19] E. Gonmwan and Y. Chali, "Transformer-based model for single documents neural summarization", In: *Proc. of the 3rd Workshop on Neural Generation and Translation*, pp. 70-79, 2019.

[20] S. Xu, H. Li, P. Yuan, Y. Wu, X. He, and B. Zhou, "Self-attention guided copy mechanism for abstractive summarization", In: *Proc. of the 58th annual meeting of the association for computational linguistics*, pp.1355-1362, 2020.

[21] J. Jiang, H. Zhang, C. Dai, Q. Zhao, H. Feng, Z. Ji, and I. Ganchev, "Enhancements of attention-based bidirectional LSTM for hybrid automatic text summarization", *IEEE Access*, Vol. 9, pp. 123660-123671, 2021.

[22] H. Shahmohammadi, M. Dezfoulian, and M. Mansoorizadeh, "Paraphrase detection using LSTM networks and handcrafted features", *Multimedia Tools and Applications*, Vol. 80, No. 4, pp. 6479-6492, 2021.