The T5-LSTM-RNN: A Text Summarization Model for the Behavioral Biology Literature

Shivangi Chaurasia[a], Debalay Dasgupta[b],

[a]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, shivangi.chaurasia2019@vitstudent.ac.in
[b]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, debalay.dasgupta2019@vitstudent.ac.in

Abstract

Behavioral biology is one of the crucial and trending topics these days which needs proper attention by scholars for rapid development in this field. Therefore, the purpose of this paper is to ease the process of collecting all kinds of relevant and vital data available on the internet regarding this topic in different forms of media, such as news articles, research papers, and YouTube lecture videos, all into one place into a single document in a proper summarized form. For proper training of the LSTM model, the lengthy video and journal datasets are pre-processed using the T5 transformer model to generate a uniform training dataset. So, in this work, a comprehensive approach is proposed based on an abstractive form of text summarization using the seq2seq encoder-decoder model combined with a stacked LSTM layer with an attention mechanism and a T5 transformer model pre-processor. Therefore, a proper hybrid model, T5LSTM-RNN, is implemented to generate the summarized data.

Keywords: Seq2Seq; Text Summarization; Abstractive Summary; T5; Transformer; Behavior Biology; LSTM; RNN

1. Introduction

The massive amount of textual data produced in the biomedical area is a constant source of difficulty, driving researchers to develop novel domain-specific text processing and summarizing algorithms. The text summary is a method of reducing a document into a more manageable size while retaining vital information. Extractive and abstractive document summarization are the two forms of text document summarization [1]. Scholars can use biomedical literature to assess the most recent developments in a given field of study, generate and validate novel ideas, conduct studies, and interpret the results. These textual resources are rapidly increasing, making it more difficult to extract and manage information from massive volumes effectively [2]. Therefore, it is critical in academia and industry to develop automated systems that simplify the tedious duties of information extraction and knowledge discovery from textual resources. In recent decades, extractive and abstractive automatic biomedical text summarizing approaches have been extensively researched [2,3,4] to provide biomedical researchers with tools to help them deal with vast amounts of information hidden in various types of resources such as news articles, videos and even journals.

Behavioral biologists are focused on the behavior of the whole organism, which is separate from both internal electron activity and the continuous genomic processes going on inside the organism. The whole unit of Life, along with the basic structure, allows parts to engage in metabolism growth and reproduction which also requires functional information processing and synchronization. However, ontologically distinguishable behavioral patterns are also involved, which is strongly supported by the distinction between Life and Mind explicit by ToK[5]. Academic articles follow an inherent structure depending upon the source. Methods of abstractive summarization produce highly reliable information and a less redundant summary [6].

The paper initially introduces some background knowledge regrading various existing text summarizations approaches in section 2 and how structured based Abstractive text summarization, will give better results as it uses ontological principles as a baseline [1,7]. The individual components of the proposed model and the methodology is described in section 3. The entire experimental setup, including how raw data is preprocessed and the comparison of ROUGE F1 scores of the T5LSTM-RNN model with the seq2seq RNN baseline model is discussed in section 4. The conclusion highlights some of the findings and makes recommendations for further research in section 5.

2. Related Studies

This chapter examines related research on several types of existing text summary approaches, as well as the need for distinct text summarization models, both pretrained and hybrid models, on various sources of content. Notable work has been done on the subject of text summarization using NLP-related domains. However, in the context of behavioral biology, summarizing large textual materials, long-duration movies, and news articles is still a relatively new research area.

Extractive text summarization is a procedure that involves awarding a value to phrases based on a bunch of factors. This summary incorporates elements at the lexical items level. For extractive text summarization, there are two basic categories: unsupervised learning and supervised learning. Summaries are difficult to assess (either automatically or manually). The inability to define a standard against which the findings of the systems to be evaluated may be compared is the fundamental problem with evaluation [8]. Furthermore, defining an appropriate summary is challenging since the system may produce a superior summary that differs from any human summary used as a rough approximation to the right output. Current trends and problems in biological text summarizing are also discussed in recent survey publications [8,9,13]. Domain-specific- study is performed to see how useful idea frequency is as a single variable in extractive text summarization for finding relevant sentences [4]. Three domain experts hand-extracted summaries from 24 biology editorials as the first step in the evaluation. A number of automatic summarizers were then used to summarize the contents. The ROUGE score was selected as the main parameter for deciding the best model [3,4]

The output sequence is produced in two steps, based on input sequence, using a unique framework that focuses on a single-document multi-sentence summary. BERT model was used to encode input sequence in context values for the encoder. The model includes two steps for decoder part: the first is to create a new output sequence using a Transformer-based decoder [1,8,10]. Most of the words in the sequence is masked and provided to BERT in next stage; By combining input sequence with draught representation generated by BERT, a Transformer-based decoder estimates the optimized word to every masked location [11]. PEGASUS tool is proposed for huge text documents with the latest self-supervised objective [7].

A sequence-to-sequence problem is widely used to model abstractive summarization. Nevertheless, using constrained supervised summarizing data to train large SEQ2SEQ-based summarization models is challenging. Three sequence-to-sequence pretraining (or STEP) objectives that allow us to pretrain a SEQ2SEQ-based completely abstractive summarization of unlabeled text are present. The basic idea is that, given artificially generated textual material from a report, a model is pretrained to restore the original document [1]. To support variable-sized inputs, the T5 model, which is built on abstractive summarization transformer architectural concepts, employs a stack of self-attention layers rather than typical RNNs and CNNs.

PEGASUS is another example of an extensively used pretrained model based on abstractive summarization principles. Several gap-sentence analysis strategies are also examined, demonstrating that selecting the principal sentence is the optimum strategy [7]. Furthermore, in a video-to-text summarization approach, a deep convolutional neural network model (CNN) based architecture is implemented. The audio from the video file is then combined using keyword-based sentence extraction. The transformer Sum model is proposed next, which simplifies the process of long text summarization by utilizing pretrained transformers such as BART and PEGASUS. [12]

3. Proposed model design

The different sources of information based on behavior biology text summarization: news articles, research journals, and YouTube videos in biomedical. Since biomedical is a vast field, this paper mainly focuses on one of its subparts, behavior biology. So, in this chapter, the model proposed focuses on extracting information from three different sources of media and extracting a summary of the same using different abstractive text summarization models and then combining all the collected summaries in a document to make it easier for scholars to access those conveniently without any wastage of time and energy in going through extensive articles and videos related to behavior biology.

3.1 T5 model for pre-processing research articles and journals

T5 model will summarize the long video transcripts and research journals before sending them to the LSTM model for a better abstractive summary. Detailed reasoning is provided behind T5 model selection in section 4.3. The T5 transformer model produced excellent results when processed over CNNDM, MSMO and XSUM, with over 42 ROUGE and 43 BLEU scores on the MSMO dataset [10]. The T5 model uses stacked self-attention layers and is composed of encoder-decoder layers followed by a feed-forward network. The following figure shows the detailed architecture [10].

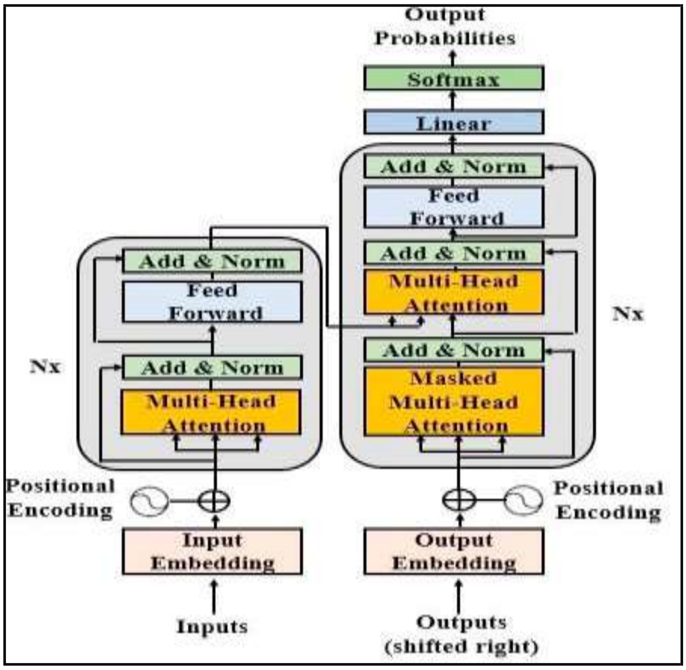

Figure 1: Transformer Model Architecture [10]

3.2 Sequence-to-Sequence RNN baseline model

A training process that transforms sequences from one domain to sequences from another is known as the sequence to sequence learning. It is a common strategy used in machine translation and other circumstances where free-form sentences are required. The input sequence in this work is the large sequence of words from the original document, and the output is a sequence of words that transmit the same idea but with fewer words, which produces the summary. The conversion employs vector mapping and prediction algorithms. Encoder and decoder are two components of a seq2seq model. The architecture of this model is designed to allow it to process information that is not limited in length. Hidden state in a simple RNN is computed using the following equation [13]:

$$h_t = \phi(W_{xh}x_t + W_{hh}h_t{-}1 + b_h) \quad (1)$$

Here, the activation function is denoted by ϕ. The secret states in encoder is denoted by h_t and the previous hidden state by h_{t-1}. The hidden states are linked, represented by W_{hh} weight matrix. The input is linked with the hidden states and is represented by W_{xh} matrix. b_h represents bias vector to h_t and x_t represents input sequences

To estimate output units, the following equation can be implemented [13]:

$$z_t = \sigma(W_{hz}h_t + b_z)$$

where σ denotes the output layer's activation function, b_z substitutes as a bias vector to the output layer, and W_{hz} represents a weight matrix from the hidden layer to the output layer

a. **Encoder Model:** This model encodes or changes input phrases while also providing feedback at each level. This feedback can be an internal state (hidden or cell) using the LSTM layer. Encoder models extract the most important information from input phrases while maintaining the context. Contextual information will be captured in the encoder's input without affecting the meaning of the input sequence. The encoder model's outputs are then passed into the decoder model.

b. **Decoder Model:** The decoder model is used to predict target sentences word by word. The target sentence is then utilised to deduct the next word before being transmitted to the prediction layer. The model uses delimiters like 'sos' and 'eos' to forecast the next word and define the sentence's ending. 'sos' is added at the beginning of the training process to anticipate the following word (the decoder target data). The delimiter chosen is then used as input data for the next timestep, allowing the next word to be predicted.

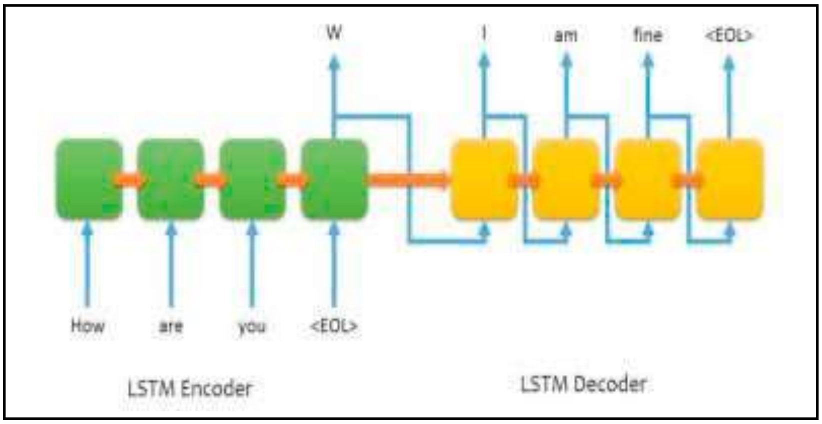

Figure 2: Sequence to sequence Model (LSTM encoder-decoder) [14]

3.3 Stacked LSTM Layer

When looking at the general Long-term short memory (LSTM) layer, it can be seen that the LSTM layer either creates output for each input or builds a feature vector, which is then used by dense neural network layers for classification tasks utilizing SoftMax layers. Because the output size is independent of the input size and both are sequences, this model technique cannot be employed in the sequence to sequence problem. The encoder-decoder network paradigm was created to address this specific challenge. The basic encoder and decoder features of a recurrent neural network (RNN) model can be comparable to those found in existing simple RNNs, LSTMS, and GRUs.

3.4 Attention Mechanism

During the creation of the sequence2sequence framework, the attention mechanism was primarily implemented to overcome the limitations of the traditional encoder-decoder paradigm. So, rather than focusing on the entire input sequence, the attention mechanism seeks to focus on specific sequences from input to anticipate a word. This method looks to fix the issue and is quite similar to the human method. Attention mechanisms can be classified into two parts, Global attention utilizes every hidden state from each encoder step to build a context vector. Another is local attention, which generates the context vector using only a handful of the concealed states.

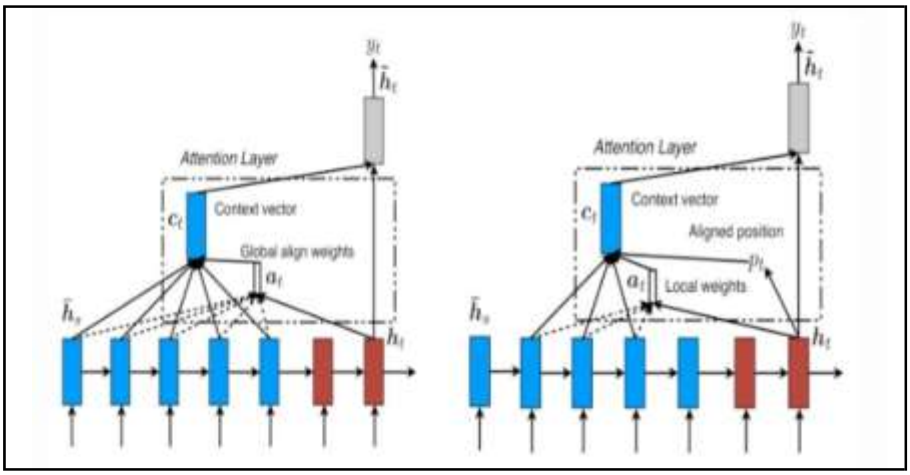

Figure 3: Global (L) & Local (R) Attention [15]

3.5 Proposed T5LSTM-RNN hybrid model

The novel proposed architecture is a hybrid model combining Sequence to Sequence RNN (section 3.2) with stacked LSTM layer (section 3.3), and Attention mechanism (section 3.4) along with the T5 transformer model (section 3.1), which outperforms other pretrained models where number of grammatical mistakes per unit length (50 words) is the chosen parameter for comparison (section 4.3)The provided input text from long video transcripts and journals automatically gets processed by T5 to generate input for the LSTM model. If the input length is equivalent to that of a news article, it gets directly processed by the attention based seq2seq RNN LSTM component. The process of training set generation along with handling of various sources of input are depicted in the following architecture diagram.

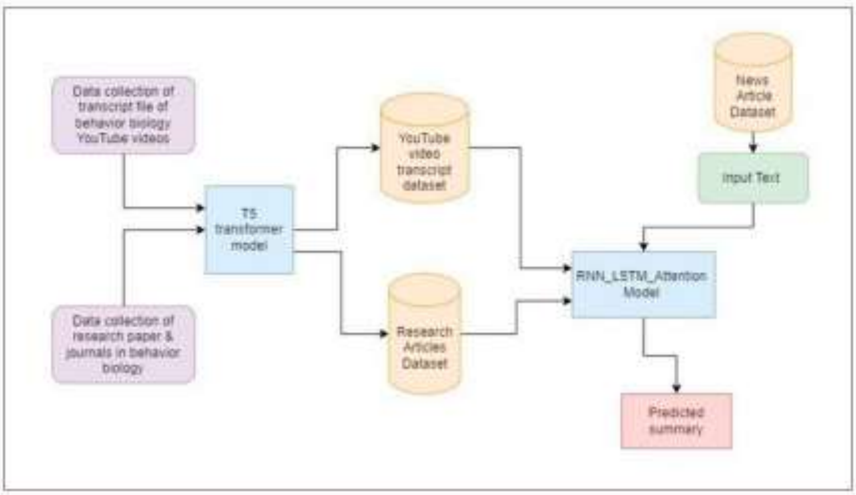

Figure 4: Proposed hybrid model

4. Experimental Results

This section begins with a detailed description of the various formats of datasets and experimental setup used for the experiments. The T5LSTM-RNN model is trained on the dataset generated through T5 transformer and on news article dataset. Finally, the proposed hybrid model is evaluated on a test dataset generating predicted summary

4.1 Experimental Setup

There are mainly three sources for data collection used for this experiment. The dataset contains over 4500 samples of text-summary pairs which includes transcripts of long video-based lectures, writings from various research journals and news articles. The videos were used from the famous playlist, namely, Lecture Collection | Human Behavioral Biology by Stanford University, which has over 2.5 million views. Transcripts of all the videos in the playlist were generated and stored in a file using the YouTube-transcript-API, a famously used python API to extract transcript for a given YouTube video. The research journals were mainly focused on Behavioral Biology topics and were all Elsevier publications. While extracting text from the journals, it was made sure that the process got rid of the Reference section as it can plague the whole model training process. Inshorts dataset was used for training the model over several articles. It contains source, time, date, headlines and short summaries of news from around the web.

To carry out the whole experiment a system consisting of eight vCPUs and 30 GB RAM with M4000 8 GB virtual ram is used. Encoder-decoder seq2seq RNN model is implemented based on the TensorFlow 2.8.0 framework. The video transcripts and the journal extracts are first summarized using the T5 model as it turned out to be the best among other pretrained models (section 4.3) The network model used is the Seq2Seq RNN text summarization model based on stacked LSTM layers with an Attention mechanism (section 3). The complete training set is pre-processed, starting with padding start and end tokens. Then regex patterns and contraction mapping have been used to remove shortened words and special characters to finally stem the input words to their root words. After the cleaning process, duplicate words are filtered and sorted accordingly to store the total number of input and target words. TensorFlow's

4.2 Experimental results and evaluation

The proposed LSTM model is trained on various news article sources; hence the model quite underperformed on extensive journal papers and YouTube videos of an average one-hour length. To overcome this constraint, a comprehensive comparison between the best pretrained abstractive text summarization models was performed. T5, Hugging face and Bart transformer model were used to process 20 Behavioral Biology papers published by Elsevier publication and the transcripts of all the 25 video lectures from Stanford University's behavioral Biology YouTube playlist. Length of the first-hand summary of the resources generated and the grammatical correctness were the two important metrics taken into consideration. For detecting grammar errors and spelling mistakes, language_python_tool was used, which is an open-source grammar tool. The results were plotted, and the average was calculated, which resulted in the selection of the T5 model. The flowchart is shown in the figure

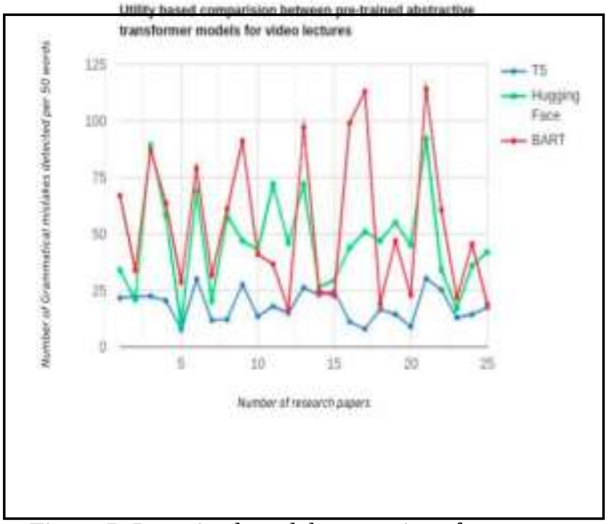

Figure 5: Pretrained model comparison for papers

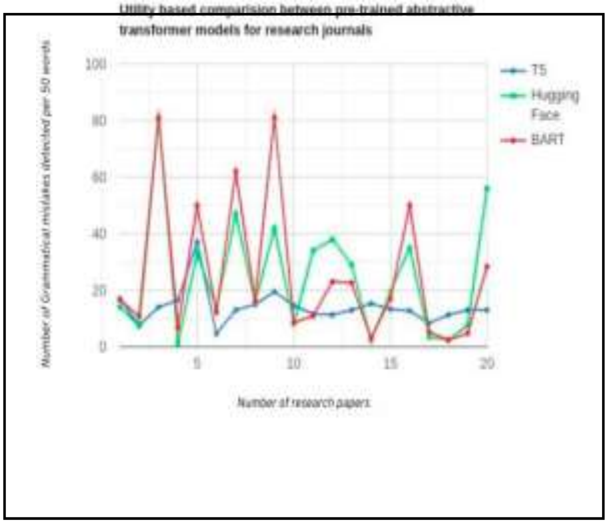

Figure 6: Pretrained model comparison for videos

Initially when the proposed hybrid model was trained with epoch value of 50, at that point loss starts diverging after 40 epochs. As the two plots keep diverging to a greater extent after the epoch value of 20, this denotes that model slowly started overfitting the data. So, it could be concluded that model trained till epoch value of 20 would deliver proper results.

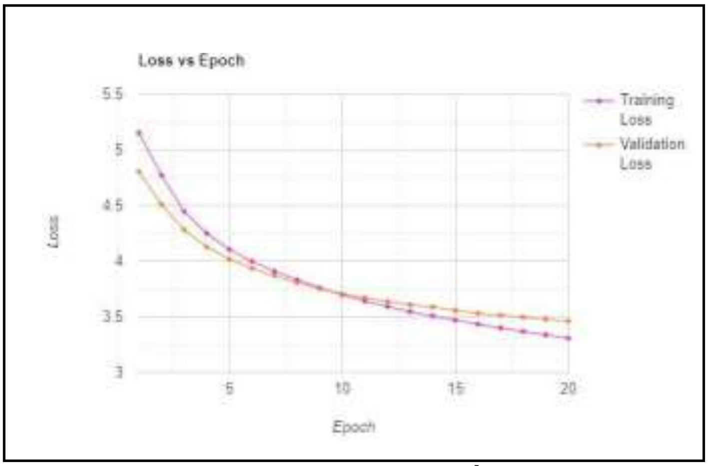

Figure 7: Loss v/s Epoch

ROUGE metric is used for comparative analysis between the proposed hybrid model and the seq2seq RNN baseline model. ROUGE-1 and ROUGE-2 are calculated using unigram and bigram overlaps respectively while ROUGE-L is based on longest common sub-sequences [16,17].

Table 1 represents the ROUGE F1 scores of both the T5LSTM-RNN and the baseline Seq2Seq RNN model.

Model	ROUGE-1 F1	ROUGE-2 F1	ROUGE-L F1
T5LSTM-RNN with Attention mechanism	0.2834	0.1132	0.2629
Baseline Seq2Seq RNN model	0.2541	0.1066	0.2354

Table 1: ROUGE scores of both the models

Table 2 represents the detailed description of an output generated by the proposed T5LSTM-RNN model on video source, research paper and news article.

Title	Input Text	Summary Generated by System / Output	Summary Length
YouTube Video- Behavioral Genetics II **Duration -** 1:32:44 	14751	metro dogs are dogs that live in the subway system in moscow and russian but they live in there. they are being selected for the exact same traits that went into being wolves. they've got a different shaped tail it's just part of the package if you were going to have this canine like thing and be selected for being scared of humans functioning and packed having a fairly aggressive temperament.	71
Paper Title- The Development of Food Search Behavior by Rats: The Effects of Hippocampal Damage and Haloperidol[18] **Authors -** ROBERT D. OADES 1 AND ROBERT L. ISAACSON z'3	3569	rats were required to locate four pellets of food located in an enclosed arena. three groups of animals were studied in testing sessions. animals with hippocampal lesions visited more nonfood holes than control animals.	34
News Article: **Source:** animal behavior news science daily (April 2022) 	697	researchers create first-of-its-kind animal model of interoception. interoception refers to the ability to sense the internal state of one's body. dysfunctions in interoception are associated with anxiety, depression, and Alzheimer's disease.	31

Table 2: Summary obtained by proposed model

5. CONCLUSION

This research presented a hybrid text summarization model in behavior biology, which solves the issues faced by scholars in accessing the complete information regarding the biomedical domain scattered over the internet. The T5 component of the T5LSTM-RNN model processes input text only if it exceeds 1000 words; otherwise, it is sent directly to the seq2seq attention model. A significant drawback of the model is that it only produces a high-level overview of the document. In some circumstances, more thorough information about the critical elements that cannot be covered in summary is desirable. As part of future work, it is also proposed to incorporate a multi-task learning strategy that unifies phrase choice and paragraph prediction methods

REFERENCES

1. Zou, Y., Zhang, X., Lu, W., Wei, F., & Zhou, M. (2020). Pretraining for abstractive document summarization by reinstating source text. arXiv preprint arXiv:2004.01853.
2. Moradi, M., & Ghadiri, N. (2019). Text summarization in the biomedical domain. arXiv preprint arXiv:1908.02285.
3. Luo, C. Y., Cheng, S. Y., Xu, H., & Li, P. (2022). Human behavior recognition model based on improved EfficientNet. Procedia Computer Science, 199, 369-376.
4. Reeve, L. H., Han, H., & Brooks, A. D. (2007). The use of domain-specific concepts in biomedical text summarization. Information Processing & Management, 43(6), 1765-1776.
5. Henriques, G., & Michalski, J. (2020). Defining Behavior and its Relationship to the Science of Psychology. Integrative Psychological and Behavioral Science, 54(2), 328-353.
6. Mohan, M. J., Sunitha, C., Ganesh, A., & Jaya, A. (2016). A study on ontology based abstractive summarization. Procedia Computer Science, 87, 32-37.
7. Zhang, J., Zhao, Y., Saleh, M., & Liu, P. (2020, November). Pegasus: Pretraining with extracted gap- sentences for abstractive summarization. In International Conference on Machine Learning (pp. 11328- 11339). PMLR.
8. Moratanch, N., & Chitrakala, S. (2017, January). A survey on extractive text summarization. In 2017 international conference on computer, communication and signal processing (ICCCSP) (pp. 1-6). IEEE.
9. Gayo-Avello, D., Álvarez-Gutiérrez, D., & Gayo-Avello, J. (2004, January). Naive algorithms for keyphrase extraction and text summarization from a single document inspired by the protein biosynthesis process. In International Workshop on Biologically Inspired Approaches to Advanced Information Technology (pp. 440-455). Springer, Berlin, Heidelberg.
10. Bohra, M., Dadure, P., & Pakray, P. (2022). Comparative analysis of T5 model for abstractive text summarization on different datasets.
11. Zhang, H., Xu, J., & Wang, J. (2019). Pretraining-based natural language generation for text summarization. arXiv preprint arXiv:1902.09243.
12. Housen, H. T. Lecture2Notes: Summarizing Lecture Videos by Classifying Slides and Analyzing Text.
13. Chen, Y. Y., Lv, Y., Li, Z., & Wang, F. Y. (2016, November). Long short-term memory model for traffic congestion prediction with online open data. In 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC) (pp. 132-137). IEEE.
14. Panesar, K., & Mudikanwi, L. (2020). Chatterbot implementation using transfer learning and LSTM encoder-decoder architecture. International Journal, 8(5).
15. Kanwal, N. (2020). Dilated Convolution Networks for Classification of ICD-9 based Clinical Summaries (Doctoral dissertation, Politecnico di Torino).
16. Lin, C. Y. (2004, July). Rouge: A package for automatic evaluation of summaries. In Text summarization branches out (pp. 74-81).
17. Mridha, M. F., Lima, A. A., Nur, K., Das, S. C., Hasan, M., & Kabir, M. M. (2021). A survey of automatic text summarization: Progress, process and challenges. IEEE Access, 9, 156043-156070.
18. Oades, R. D., & Isaacson, R. L. (1978). The development of food search behavior by rats: the effects of hippocampal damage and haloperidol. Behavioral biology, 24(3), 327-337.