Abstractive Text Summarization By Using Deep Learning Models

Soma Shrenika[a], GouriPriyaRamini[a], LunsavathBhagavathiDevi[a], B Geetavani[b]

[a]Under Graduate Student, Computer Science and Engineering, Institute of Aeronautical Engineering, Hyderabad, India
[b]Assistant Professor, Computer Science and Engineering, Institute of Aeronautical Engineering, Hyderabad, India.

Abstract: Knowledge is perpetual, a person's life time isn't enough to absorb the whole knowledge of the universe. We all are homo sapiens who believe in living beyond the nihilism. We try to seek more and more potentiality in our lives by seeking knowledge. This knowledge these days is stored in various formats in huge repositories mostly in the form of documents, sheets, photos, videos. One finds it difficult to comprehend this whole lot of information. There by, here comes the need of text summarization. Summarization of documents, text, data is the vital part and a preliminary step in any field whether it is business, social, art, or software. By using machine learning algorithms, the notion of text summarization can be achieved with ease. In this paper we summarize the data as per our requirement i.e., it can be based on output. To achieve this objective we are going to use abstractive summarization on single documents. In this process, we perform abstractive text summarization by using Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU) and Transformers.
Keywords: Text Summarization, Abstractive Text Summarization, Long Short Term Memory, Gated Recurrent Unit, Transformers

1. Introduction

Text summarization is nothing but obtaining only required information from huge data. Since manual text summarization is a time costly, for the most part a relentless task. Programmed text summarization establishes a solid inspiration for additional examination. Current endeavors in programmed text summarization basically center on summing up single records (for example news, articles, logical papers, climate figures, and etc) also, multi reports (for example news from various sources, client surveys, messages and etc), lessening the size of the underlying content while simultaneously protecting key enlightening components and the importance of substance. Two fundamental ways to deal with programmed text summarization have been ac-counted for in the significant writing; extractive and abstractive.

Abstractive text summarization is a really difficult undertaking, it looks like human-composed rundowns, like that might restrain reworded decisions instead expressions including current term(for example abstract, expressions words which are not exposed in those first content), in this manner improving the created rundown as far as union, coherence or repetition . As computer does not understand human language and the idea related with it, summarization become a difficult work. To crush this, many simulated intelligence models like machine learning and deep learning are used. These models are set up in a manner of comprise to include huge information. The principle commitment related to this is to develop an intellectual text summarization strategy.

2. Literature Survey

SukirtiVerma et.al, [1], developed a system which upgrades include vector and esteems are attach against each sentence to create a score. According to decreasing score value the sentences are arranged. The abstract applicable sentence is the main sentence used in this arranged rundown and it is collected as a feature of the subclass of sentences will frame the summary. The sentence we select is the main sentence having most interesting likeness with the principal sentence, which is selected stringently from the top of the arranged shorted list. This interaction is repetitive and gradually iterate to choose more sentences summary limit to reach the user deter-mined limit. Then they are re-organized according to the pattern surfacing in the orignal text. This disorganized sentences will deliver an appraised outline instead of a group.

ManivannanKaliapan et.al, [2], proposed a subsequent method to handling the archives by utilizing the Naïve Bayesian approach. Based on the results the sentences are organized. First the high score sentence will be generated, then after the other sentence, etc. The Navie Bayesian process will choose sentences for each record to produce results. They consider sentences previously to connate decision chosen of following record eventually. To defeat some issues timestamp strategy is carried out. The timestamp technique is carried out by allotting a worth to each pronouncement of the document. It depends on sequential situation of report. When the determination are chosen they are promptly organized in the climbing request which dependent on checksum. The arranged view assists with accomplishing the rundown, this completes a consistent looking encapsulation. This complete number of decisions in the abstract is addressed by the squeeze amount.

3. Data Description

For the purpose of single document text summarization we have used "News Summary" dataset from Kaggle. The dataset consists of two files "news_summary.csv" and "new_summary_more.csv". Both of them are used in dif-ferent methods. The news_summary file contains 4516 documents of data, where each of it contains 6 columns namely; author, date, headline, readmore, text, ctext. The news_summary_more file contains 98360 documents which contain 2 columns, namely headlines and text.

3.1 Deep Learning

It is a part of ANN. It consists of various architectonics which deal with various human computer interaction problems and consists of networks that are based on graphs, convolutions, etc. The problem with linear perceptron is that, it does not work as expected for certain problems. This led to the introduction of multiple layers, which is known to be deep learning. It increases the precision of computation. These algorithms can be implemented with unsupervised learning. Neural Networks (NN) are used for natural language processing from more than 20 years. These neural networks mimic the neurons present inside the brain. It consists 3 layers in general. It can contain more than three layers, depending on the requirement. These convert the input into righteous output.

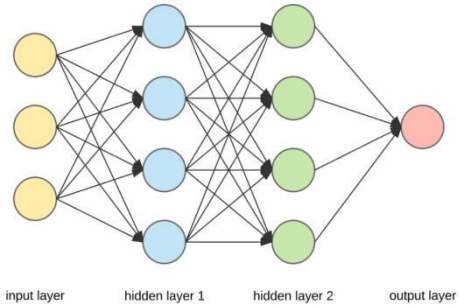


Figure1 Neural Network

3.2 Recurrent Neural Network

RNN have an internal memory unit. These memory units are very useful in models that deals with data which is in a sequential form. In other network architectures, the decision at a point only depends on its present input and output is produced. Whereas, in RNN previous output is also taken as input for next computation. The concept of back propagation is used in RNN. It evaluates the mathematical partial derivation of the error wrt weight, then gradient descent is used to reduce the error. Back Propagation Through Time (BPTT) is used to implement RNN. It is used with unrolled RNN. It computes the error and changes the weights for every step.The below figure shows an unrolled RNN.Here, x_t represents the input for time t, ranging from 0 to time t. h_t is the output for every value of t. A is used to represent neural network.

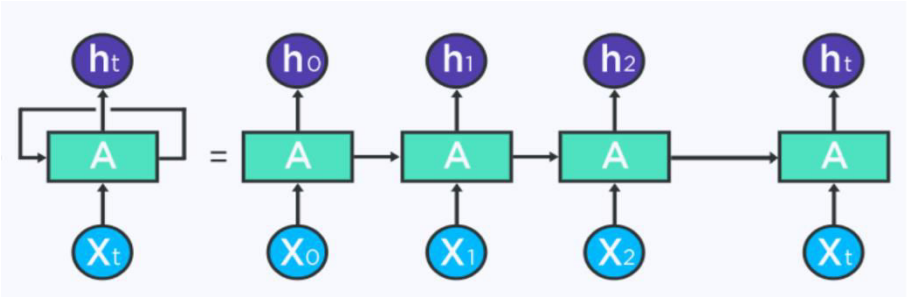


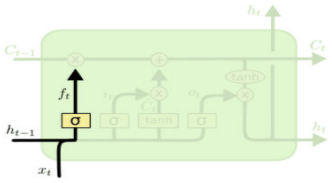
Figure 2 Unrolled RNN

But there are 2 issues with RNN. If the partial derivative values are more than huge, then the values of gradient will increase, which leads to exploding gradients. Similarly, if the partial derivatives are very less, the values of

the gradient will fall down, which eventually will result in vanishing gradients. To retain the inputs for larger time, LSTM networks are used.

3.3 Long Short Term Memory (LSTM)

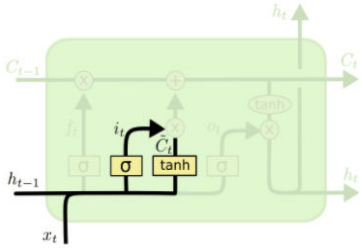
The LSTM architecture consists 4 neural network layers and has 3 gates to operate on the state of a cell. The first gate is the "forget gate". It is used to make a decision by using a sigmoid function whether to preserve the information or discard it by rendering a value of 0 or 1 by h_{t-1} and x_t.



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

Figure 3 Forget Gate

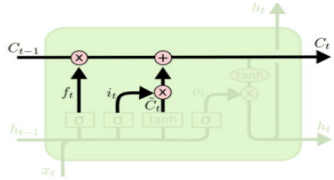
Then, the process involves in understanding storage of information in a cell state. The "input gate" consists of a sigmoid function. Another layer calculates tanh and produces a vector. Both the above layers are then used to update the state.



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

Figure 4 Input Gate

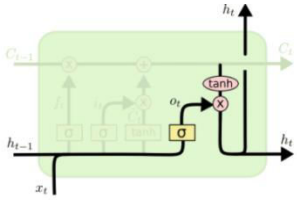
Updation is done by multiplying $f_t$with cell state c_{t-1} and then adding i_t *C_t



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Figure 5 New Candidate Values

The last gate is the "output gate". It is used to obtain the output. For calculating this we use a tanh layer and a sigmoid layer.



$$o_t = \sigma\left(W_o \; [h_{t-1}, x_t] \; + \; b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Figure 6 Output Gate

3.4 Gated Recurrent Unit (GRU)

It is similar to LSTM but has only two gates and is faster. The first gate is "update gate", it is used to obtain the amount of information to be passed forward. This is used to remove the vanishing gradient problem. It is obtained by using a sigmoid function. Here, the values obtained will be in between 0 and 1.

$$z_t = \sigma(W^{(z)}x_t + U^{(z)}h_{t-1})$$

The second gate is "reset gate". It is used to find the part of informatio to be removed.

$$r_t = \sigma(W^{(r)}x_t + U^{(r)}h_{t-1})$$

Current memory content uses the reset gate to remove irrelevant information. A tanh function is applied on sum of two elements. The first is the product of weight w and input x_t . The second one is product of the reset gate and weight of h_{t-1} by using the hadamard product.

$$h_t^{'} = \tanh(Wx_t + r_t \odot Uh_{t-1})$$

At last we need to find the information that is needed to pass forward. This is done by using the update gate. Apply hadamard product between z_t and h_{t-1} nad $1\text{-}z_t$ and $h_t^{'}$. Information from the current memory and the previous memory are obtained and added.

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot h_t^{'}$$

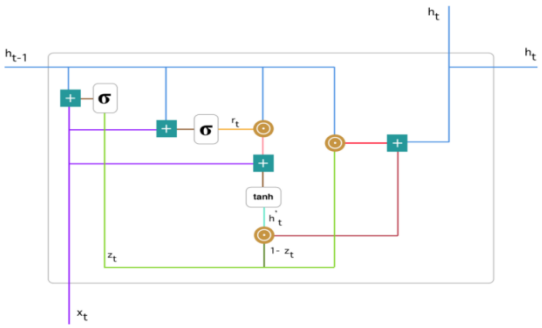


Figure 7 GRU

3.5 Transformer T5

Transformers are used to obtain parallelization. It is used to solve sequence to se-quence problems. It consists of encoders and decoders which are multiple in number and they are stacked . T5 is a text to text transfer transformer. It receives input in the form of text and output is also in the form of text. It is used for natural language pro-cessing tasks like, summarization, language conversion, etc. This model was original-ly developed by Google.

4 Methodology

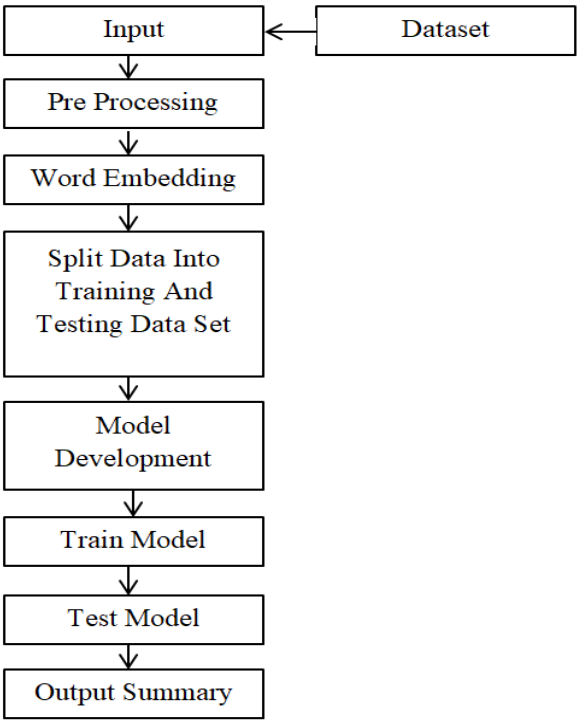


Figure 8 Block Diagram of Proposed Methodology

4.1 LSTM Implementation

- Import numpy, pandas, re, tensorflow, nltk, attention, etc packages.
- Clean the data by removing escape characters, url's, multiple spaces, email con-tent. Use the pipe method of spacy to speed up the cleaning of data.
- Obtaining information on length of text and percentage of data with the obtained length.
- Split 10% of data as test data set and remaining 90% as training data set.
- Convert the data into tokens. Find the total count of words, count of unique and rarely used words.
- Convert these tokens into vectors. Converting to vectors is converting data into integers. To make all sentence to have same length, pad zeroes upto maximum length.
- Now train the model
  - Add a layer of embedding of 200 dimensions for the text data which is in the form of sequences.
  - Add three layers of LSTM for the encoder of latent dimension value equal to 300.
  - Add a LSTM layer for the decoder of latent dimension value equal to 300.
  - Add a dense layer with an activation function. The argument value for the activation function is softmax.
- Start fitting the model with the data. Train on 88517 samples and validate on 9836 samples. 50 epochs are used for training.
- Define decoder function to infer the implementation.
- Run the model to see the final results.

4.2 GRU Implementation

- Import all the required packages and define contraction mappings.
- Clean the data by removing unwanted data like punctuation, additional spaces, etc.
- Split every line into pairs and normalize it.
- Define the encoder, decoder and the attention functions.

- The attention mechanism used is dot product of hidden states. Then these states are passed to a softmax activation function and provided as input for next hidden states.
- Teacher forcing ratio is used to speed up the process and provide the ground truth value instead of providing the predicted outputs for the decoder.
- Train it for 150000 iterations.
- Run the model to obtain results.

4.3 Tranformer T5 Implementation

- Libraries imported are:Pandas, Pytorch, PytorchUtils for Dataset and Dataloader, Transformers, T5 Model and Tokenizer.
- Clean the dataset to remove unwanted columns.
- The data is divided into 80-20 ratio for test and validation.
- Train and Validation parameters are defined and passed to the pytorchDataloadercontstruct to create train and validation data loaders.
- Define the model and optimizer that will be used for training and to update the weights of the network.
- Train the model with all the necessary parameters.
- Generate the summaries.

5 Results

Inputs and outputs for all three algorithms are provided in this section.

pope francis on tuesday called for respect for each ethnic group in speech delivered in myanmar avoiding reference to the rohingya minority community as the nation works to restore peace the healing of wounds must be priority he said the pope myanmar visit comes amid the country milit ary crackdown resulting in the rohingya refugee crisis
students of government school in uttar pradesh sambhal were seen washing dishes at in school pr emises on being approached basic shiksha adhikari virendra pratap singh said yes have also recei ved this complaint from elsewhere we are inquiring and action will be taken against those found guilty
apple india profit surged by 140 in 2017 18 to crore compared to ã¢â□â¹373 crore in the previou s fiscal the indian unit of the us based company posted 12 growth in revenue last fiscal at ã¢â□â¹ 13 crore apple share of the indian smartphone market dropped to 1 in the second quarter of 2018 according to counterpoint research
uber has launched its electric scooter service in santa monica us at 1 to unlock and then 15 cents per minute to ride it comes after uber acquired the bike sharing startup jump for reported amount of 200 million uber said it is branding the scooters with jump for the sake of consistency for its o ther personal electric vehicle services
around 80 people were injured in accidents related to kite flying during celebrations of makar san kranti in rajasthan jaipur officials said the victims included those who fell while flying kites and t hose injured by glass coated kite string officials added meanwhile around 100 birds were reporte d to be injured by between january 13 and 15

Figure 9Input for LSTM Algorithm

Predicted summary: start pope francis urges myanmar to withdraw rohingya violence end
Predicted summary: start up school students fall ill after eating mid day end
Predicted summary: start apple india profit rises to ã¢â□â¹1 crore in march quarter end
Predicted summary: start uber launches its own electric car service in 2019 end
Predicted summary: start people injured in accident in rajasthan end

Figure 10 Output for LSTM Algorithm

usbased study found brain cognitive abilities may significantly reduced mere presence smartphon e even switched off . series tests requiring concentration participants phones another room outpe rformed phones desk also kept phones pocket bag researchers said .
union minister ravi shankar prasad friday said though muslims vote bjp bjp government given pr oper sanctity . we got 13 chief ministers own . ruling country . victimised muslim gentleman w orking industry service dismissed them asked .
22000 people evacuated saturday fire broke one stages tomorrowland unite music festival barcel ona spain . believed fire could triggered firework display part show . firefighters managed extin guish flame injuries reported said organisers .
official james bond twitter handle announced november 8 2019 release date new film franchise ti tled bond 25 . film written neal purvis robert wade produced michael g . wilson barbara broccoli . daniel craig reportedly return character james bond film
outgoing uttar pradesh cm akhilesh yadav sunday said our struggle continue adding samajwadi p arty party ideology . told party workers you reach people strengthen party again . samajwadi par ty devise new strategy meeting mlas party candidates added .

Figure 11 Input for GRU Algorithm

smartphone within reach reduces brain power study
muslims vote us gave sanctity prasad
22000 people evacuated fire spain music festival
release date new james bond film bond 25 announced
struggle continue akhilesh yadav

Figure 12 Output for GRU Algorithm

PM Narendra Modi on Thursday launched Ude Desh ka Aam Nagrik (UDAN) scheme for regional flight connectivity by flagging off the inaugural flight from Shimla to Delhi. Under UDAN, government will connect small towns by air with 50% plane seats' fare capped at?2,500 for a one-hour journey of 500 kilometres. UDAN will connect over 45 unserved and under-served airports.
BJP chief Amit Shah on Wednesday slammed Delhi Chief Minister Arvind Kejriwal after AAP blamed EVMs for their loss in the MCD polls. "Let Kejriwal first tell the public how he won in Delhi in 2015 with these same EVMs..." Shah said. "UPA one and two had also won elections with the same EVMs," the BJP leader added.
Sujal Kahar, a 12-year-old football enthusiast and goalkeeper from Mumbai, has been shortlisted to represent India at the global Football For Friendship (F4F) social programme. It will be held in Russia's St Petersburg in July. India will join the F4F programme for the first time and will participate alongside over 60 other countries.
Eight member nations voted against BCCI's proposal of retaining ICC's old revenue model at the Board meeting in Dubai on Wednesday. The Sri Lankan board was the only member which supported BCCI's stance on governance, while it joined others in opposing BCCI's proposal to retain the revenue structure. BCCI will lose?1,000 crore if the new revenue model is implemented.
Former AAP leader Mayank Gandhi urged Delhi CM Arvind Kejriwal in an open letter to forget his ambitions of becoming Prime Minister and focus on governing the national capital instead. "You took complete power by becoming convener as well as CM of Delhi. How much more power do you want before you start delivering alternative politics?" he added.

Figure 13 Input for Transformer T5

UDAN (Ude Desh Ka Aam Naagrik) scheme seeks to make flying more affordable for the common people. Under UDAN, 50% of seats on each flight would have capped at?2,500 per seat/hour and 50% of the seats will have remuneration of?2,500 per hour. The scheme was launched to "make air travel accessible" in regionally important cities. It will be operated by Alliance Air, the regional arm from which PM Narendra Modi handed over boarding passes to some passengers travelling via Shilal-Delhi sector).
BJP chief Amit Shah on Wednesday slammed Delhi Chief Minister Arvind Kejriwal over allegations of faulty electronic voting machines. "Let him first tell the public how (CM) won in 2015 with these same EVM'S when the BJP was at centre," he said. He further said that UPA one and two had also won elections with the same EVM, but not from the same EVMs. If EVM machines are defective then why did they win earlier? I think them should accept verdict as per his party has been communalised politics in Bengal.
BJP has said that it will not be able to take up the challenge of re-inventing the rules in the next few months. "It's time for us to make sure that we donâ€™T waste our time," said BJP leader Anil Kapoor. He added, "We are going to have to do something about this issue." The government is also planning to launch a pilot program in the next few months.
BCCI was decimated at the International Cricket Council's Board Meeting on Wednesday, where the majority voted for a change in governance and revenue model. Both the new governance structure as well revamped revenues model saw India getting walloped by 2-8 margin while Sri Lanka voted alongside it. The only country that voted alongside BCCI was SriLankam. Notably, BCCI lost the vote on "governance" changes by 1-9 percentage points while the revenue model saw India get wallone with 2-8 percentile.
AAP leader Mayank Gandhi on Wednesday slammed Delhi CM Arvind Kejriwal in an open letter. "A defeat deals...a body blow to arrogance and makes the mind ready for reflection," wrote Gandhi. "Abandon your agenda of wanting another BJP and Cong." He further advised KejRimani to give up national ambition, focus on governing Delhi well, that is his only chance at staying relevant. Stop dramases and blameing," he added while addressing Kejrinthi'll

Figure 14 Output for Transformer T5

References

1. SukirtiVerma and VagishaNidhi, Extractive Summarization Using Deep Learning, Inter-national Conference On Computational Linguistics and Intelligent Text Processing, 2017.
2. ManivannanKaliapan and Neduchelian, Ramanujam, Automatic Multi Document Text Summarization Approach Based On Naïve Bayesian Classifier Using Timestamp Strategy, The Scientific World Journal, 2016.
3. M Afzal, F Alam, KM Malik, GM Malik, Clinical Context–Aware Biomedical Text Sum-marization Using Deep Neural Network: Model Development and Validation, Journal of medical Internet research, 2020 .
4. R Kotadiya, S Bhatt, U Chauhan, Advancement of Text Summarization Us-ing Machine Learning and Deep Learning: A Review, Proceedings of First International Conference on Computing, Communications, and Cyber-Security, 2020.
5. S Abujar, AKM Masum, MS Islam, F Faisal A Bengali text generation approach in context of abstractive text summarization using rnn, Innovations in Computer Science and Engi-neering, 2020.
6. PG Magdum, S Rathi, A Survey on Deep Learning-Based Automatic Text Summariza-tion Models, Advances in Artificial Intelligence and Data Engineering, 2021.
7. N Dang, A Khanna, VR AlluguntI, TS-GAN with Policy Gradient for Text Summariza-tion, Data Analytics and Management, 2021.