Text Summarization

Nabil Mahtab[1], Dhara K.N[2], Nishant Ranjan[3], Sourav Kumar[4], Ayush Arora[5]

[1-5]Computer Science Department, Dayananda Sagar College of Engineering

---

Abstract: *Text Summarization is defined as the process of generating summary from long text documents. As the (amount) of data around us is ever-increasing, it becomes a necessity to extract important information from long-texts. There are two methods of text summarization namely: Extractive and Abstractive. Extractive method generates summary which are the subsets of actual text, without modifying any word, whereas the summary generated by abstractive methods are more human like because they are modified to add new words, which uses NLP(Natural Language Processing). The paper explores about different extractive methods and present the implementation details based on abstractive method. Eventually, the paper shed light on the various evaluation techniques along with their advantages and disadvantages.*

A. INTRODUCTION

Automatic text summarization is the process of producing [1] a short and meaningful summary while preserving key information content of the original document. In recent years, numerous approaches have been developed and applied widely in various domains. . For instance, search engines generate snippets as the previews of the documents. News websites produces short descriptions of news topics usually as headlines which helps in browsing through the important headlines

An extractive summarization technique consists of selecting vital sentences, paragraphs, etc. from the original text based documents and concatenating them into a shorter form. The significance of sentences is strongly based on statistical and linguistic features of sentences.

Whereas Abstractive summarization technique generates summary in a way humans make notes, by incorporating important ideas. It is more challenging to generate summary of this kind as new words which may not be present in text is used in summary.

The paper presents various extractive methods which are generally used for generating summary like frequency method, which uses the frequency of words as a metric to decide the selection of sentences in the summary. Other methods include ranking based approach and Graph based model. The former takes the document as input and "tokenized" it to remove the stop words and the remaining script is thought as keywords, whereas the latter approach uses graphs and tracing out paths in from the first sentences to the last one in the document with edges signifying the quality.

This paper then goes on to explore various abstractive methods, though this method is difficult to implement but it can produce better results in comparison to the extractive methods.

The RNN (Recurrent Neural Network) method uses annotated gig word which provide millions of training data to work upon. The model architecture comprised of encoder and decoder, in which the encoder converts the original document into an internal representation and the decoder is responsible for generating the output.

The other method is known as the Neural Network approach, which the propagation algorithm as a backtracking technique, given by rank net which uses Neural Network of two layers.

Transformer method is also a type of neural network architecture that works on the basis of attention mechanism.

In conclusion, though the extractive methods are easier to implement but in some cases they produce irrelevant summaries whereas the abstractive methods is still new and under research and developmental phase, it promises to produce more relevant "Human-like" summaries.

B. Background Research and related work

2. Extractive summarization

Extractive summarization works by selecting the sentences from the original document based on its relevance, which can be the most important sentence. The task of all extractive summarizer can be divided into three independent processes: 1) Construction of an intermediate representation of the input document. 2) Scoring of the sentences based on representational parameters. 3) Selecting the summary by choosing the important sentences.

2.1.1 Intermediate representation

Every model creates some intermediate representation of the raw input text, which tries to find the relevant sentences for the summary. Topic representation and indicator representation are two such methods, the first one interprets the topic discussed in the text while the latter describes every sentences as list of features like sentence length, position of sentence in the document etc.

2.1.2 Assigning Score to Sentences

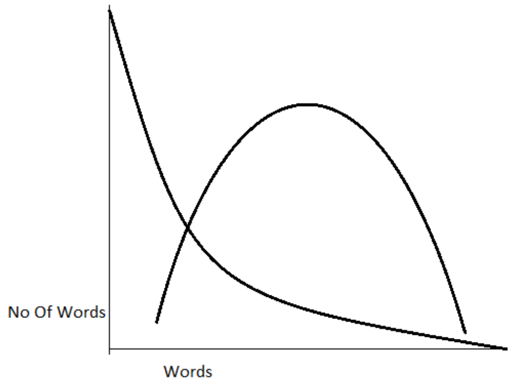
The sentences generated from [3][4] the intermediate representation are assigned a score based on different parameters like content word feature, title word feature, biased word feature etc. In the same manner sentences are also assigned score based on parameters like sentence location, sentence length, and sentence-to-to sentence cohesion.

2.1.3 Selecting sentences for summary

Finally, the most important sentences are selected to form the summary. The selection algorithm can be a simple greedy approach or it may convert it to an optimization problem. The various factors can affect the selection of sentences, for instance the type of document, the context in which summary is generated etc.

2.2.1 Frequency based Approach

H.P Luhn [6] was one of the first to come up with the automatic creation of abstraction of the given literature survey. The abstract created was almost influenced by the background of the author and may even sometime vary when it's written by the same author over time. Measuring the Significance of the text can be one way to get the summarization of the text in which the significance factor is calculated based on the analysis of the words which takes into account the frequency of word occurrence. The more often certain words are found in a sentence, the more significance may be attributed to each of these words. The words which are too common to appear in the system are considered to be the noise such as the propositional and connectors words like "is", "are" ,"and" etc. Taking this "significance factor" into account can then reflects the number of instance of most significant words within a whole sentence and then calculating the distance linearly and then the interposition of non-significant terms. All sentences may be classified in order of their importance according to this factor, and one with the highest ranking may be selected. As this method came along in 1950 when computational power was very limited and hence frequency based approach is very fast in our modern day system. Now since the computational power of the modern computers has exponentially risen, therefore this method now becomes very effective for large datasets, because the training of datasets takes very less time.



2.2.3 Ranking based approach

In this [7] approach we tend to take input as a document that is tokenized to induce the tokens of the terms. During this technique all the stop words are removed when tokenization and remaining words are thought-about as a keyword .After this frequency is calculated for every keyword further as weighted frequency so individual terms ranks are calculated that is extracted by the summarizer to search out the outline of the document .The advantage of exploitation of this approach is that the summarized text is born-again into an audio kind. This technique has additionally higher accuracy in comparison to ancient approaches. One Disadvantage of exploitation of this approach is that, we tend to take away all the stop words that in some cases could function as a keyword and it doesn't take into consideration the previous word for following foretold words.

2.2.4 Graph and semantics Method

This paper [8] explains various machine learning techniques enacting theoretic summarisation, like graph and linguistics based techniques .The various challenges one faces in text summarisation are finding vital unit sentences from a cluster of sentences and to rank these sentences from an outsized dataset associate decreed to finally build associate degree intelligent outline in an understanding format. Semantic technique is employed as a result of the oppositely applied mathematical ways sometimes don't contemplate the linguistics or context of that means. Syntactic category labelling receives all transitive verbs and then it annotates, making frames out of it. Mostly Graph based technique constructs the outline by taking the shortest path that starts with the primary sentence of the first text and ends with the last sentence .The final graph output contains the complete topics coated as sub graphs and also the intelligent or substantive sentences denoted by the quantity of edges involved with it. The disadvantage of this paper is that these processes want improvement in sentence generation and role labelling.

3. Abstractive Based Methods

3.1 Sequence To Sequence RNNs Model

The Sequence to Sequence RNNs Model is a major step forward from the Extractive Summarization Technique, which consists of identifying key sentences or passages in the source and reproducing them as summary.

Model

In this model [] we used the annotated Gig word corpus which provides us with around 3.7M training data. To optimize our loss the model has used stochastic gradient descent with mini-batches of size 32. The minimization is achieved by:

$$L = -\sum_{i=1}^{S}\sum_{t=1}^{N} \log \mathrm{P}(y\, i\, t\, |\{y\, i\, 1\, , \ldots , yi\, t - 1\}, x\, i\, ;\, \theta)$$

3.2 The Recurrent Architecture

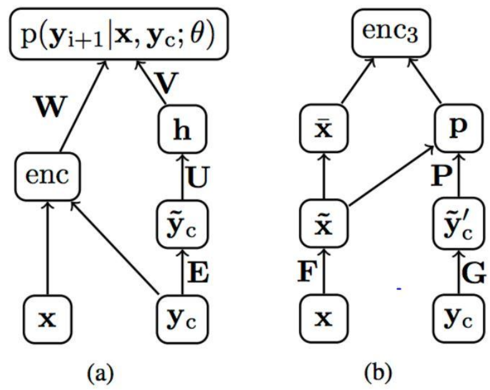
This model works with encoder and decoder in which the encoder encodes the input to an internal representation .The decoder is responsible for generating output words for the summary with encoded representation as input. The model work with the Elman RNN and LSTM-RNN and the description is as follows:

LSTM-RNNs

The LSTM-RNN encoder computes context vector for every time step t of the decoder input. The input after embedding is considered as d dimensional and the position of input and the input embedding is considered for the encoded layer.

Decoder

The hidden layers generated after feeding in the last word of the input text sentence is given as input to the decoder. The current word is given as input to the decoder for generating the next word. Hence the LSTM RNN encoder decoder model provides satisfactory result but the sequence computation inhibits parallel computation which leads to wastage of more computing power and we need to overcome this drawback with another model.



3.3 Neural Network Approach

The approach given by [8] identifies the most significant sentence in the Text. The propagation algorithm by backtracking given by Rank Net is using a neural network of two layers. Its start with tagging the data using ML and then gives features of the text in the training and testing phase. The sentence is then ranked in the document. Second approach is by using a three layer ANN network in which it learn the non-summary and the summary of the text in the training phase now the relation between the sentence are identified using the ranking of the sentence and wiping out the most rare features. Its take WikiHow articles as input and summarize the articles based on the given words

3.4 Transformers

Transformers are the type of neural network architecture that works on the basis of attention mechanism and here it is used for sequence to sequence translation. The model is trained on the Daily Mail (CNN) datasets with the batch size of around 16.

Attention

The Attention layer assigns weight to each input word such that the context representation of the input sequence is maintained and while outputting the text it parallelizes the computation and maintains the contextual aspect of the data. The Attention Softmax function is represented as follows:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d}})V$$

$$SelfAttention(X) = Attention(XW_Q, XW_K, XW_V)$$

The model is based on the stacking of the attention layers and using the encoder decoder for enhancing the computation as for large length of text input the complexity of the model is the quadratic of the length of the input and hence for longer input as in our case of summarization we have to enhance the process by using encoder-decoder/CNNS with the attention layer instead of using only the attention layer. The accuracy and the efficiency of the model have to be maintained and is a major trade-off depending on the dataset. The evaluation of the model is done with the help of the ROGUE scores which provides us with better result for the lightweight convolutional model.

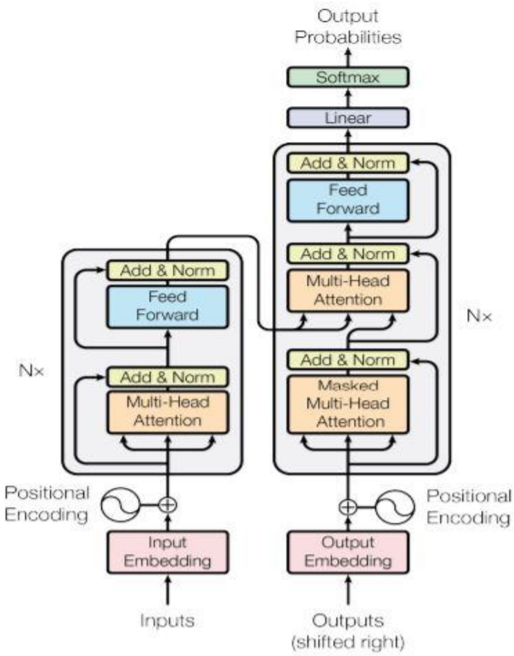


Figure 1: The Transformer - model architecture.

4. Evaluation Metrics

The task of declaring [12][13] a summary as an accurate in all context is a difficult one, but there are some indicators which can be used. For instance the most commonly used approach is human annotators, which assign a score to each sentences in summary from a predefined scale. The quality of summary can also be determined by intrinsic content evaluation, in which the summary generated by model is evaluated against an ideal summary. The text quality measure [14][15] includes metrics like grammatical correction, cohesion and sentence structure, also the generated summary should not contain irrelevant information like data redundancy.

A more rigorous method includes – Rouge method, which is an automatic evaluation technique, which are based on the similarity of n-grams.

Conclusion

The various Abstractive and Extractive methods examined in the survey paper answers few of the questions on these methods but can it be possible to combine the two approaches (extractive and abstractive)? The frequency based approach which is considered as the base for all text summarization have too much flaw in it but still can be considered as a good starting point to this approach. Abstractive summarization can be made using the decoder network which can be used to furtive the centre of the cluster into natural Language. The pointer and the generator network that gets the best of both worlds by combining both of them. The Quick thought vector which is a development of a Thought skip approach can improve the performance and reduce the training time of the model.

REFERENCES

[1] N.Moratanch, S.Chitrakala, Anna University, CEG, Chennai, IEEE International Conference on Computer, Communication, and Signal Processing (ICCCSP-2017)

[2] Deepali K. Gaikwad and C.Namrata Mahender, A review paper on text summarization, international journal of advanced research in computer and communication engineering

[3]F. Kiyomarsi and F. R. Esfahani, "Optimizing Persian text summarization based on fuzzy logic approach," in 2011 International Conference on Intelligent Building and Management, 2011.

[4] F. Chen, K. Han, and G. Chen, "An approach to sentence-selection based text summarization," in TENCON'02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering, vol. 1. IEEE, 2002, pp. 489-493.

[5] Y. Sankarasubramaniam, K. Ramanathan, and S. Ghosh, "Text summarization Using Wikipedia," Information Processing & Management,vol. 50, no. 3, pp. 443-461, 2014.

[6] The automatic creation of literature abstract H.P.Luhn

[7] 1 Text Summarization: An Essential Study [Prabhudas Janjanam and CH Pradeep Reddy

[8] Extractive Text Summarization Using Sentence Ranking[J.N.Madhuri and Ganesh Kumar.R]

[9] Ramesh Nallapati, Bowen Zhou, Caglar Gulcehre, Bing Xiang, et al. Abstractive text summarization using sequence-to-sequence rnns and beyond. arXiv preprint arXiv:1602.06023,2016.

[10] [18] K. M. Svore, L. Vanderwende, and C. J. Burges, "Enhancing singledocument summarization by combining ranknet and third-party sources." in EMNLP-CoNLL, 2007, pp. 448-457.

[11] CS224N Project Report Faster Transformers for Text Summarization(Amaury Sabran asabran@stanford.edu Alexandre Matton alexmt@stanford.edu

[12] Berry, M.W.—Dumais, S. T.—O'Brien, G.W.: Using Linear Algebra for Intelligent IR. In SIAM Review, Vol. 37, 1995, No. 4.

[13] Boguraev, B.—Kennedy, C.: Salience-Based Content Characterization of Text Documents. In I. Mani and M. T.

Maybury, eds., Advances in Automatic Text Summarization, the MIT Press, 1999.

[14] Brin, S.—Page, L.: The Anatomy of a Large-Scale Hypertextual Web Search Engine. In Computer Networks and ISDN Systems, Vol. 30, 1998, pp. 1–7

[15] Ding, Ch.: A Probabilistic Model for Latent Semantic Indexing. In Journal of the American Society for Information Science and Technology, Vol. 56, 2005, No. 6, pp. 597–608.

[16] Dumais, S. T.: Improving the Retrieval of Information from External Sources. In Behavior Research Methods, Instruments & Computers, Vol. 23, 1991, No. 2, pp. 229–236.