

Automatic Text Summarization, Types and Challenges- A Review

Sanah Sayyed, Assistant professor, Department of C.S., Shivchhatrapati college Aurangabad, Maharashtra, India, Sanahsayyed92@gmail.com

Abstract Internet produces data incrementally in today's busy life everyone wants the things easily, accurately. Absolutely human produces the summary of document accurately without any mistake, but it is time consuming task, we want such system which produces summary of documents within less time, only Automatic Text Summarization is only approach which compress and minimize the original text document by using algorithm. This paper presents the automatic text summarization, types and features of automatic text summarization, this paper outlines the review on extractive and abstractive techniques, workflow diagram of automatic text summarization and challenges.

Keywords — Automatic text summarization, abstractive text summarization, extractive text summarization, feature extraction, preprocessing, summary.

I. INTRODUCTION

Now a day's tremendous amount of data has been increased on web, data in text, image, video format. It is tedious task for human being to find the significant information, when user searches the document on the web; he must go through from different result display by the web, which leads to wastage of time and effort. We need such system which produces summary of document; Text summarization is the only application of natural language processing (NLP), which produces summary of document of relevant content without modifying the actual meaning of documents. Summary will help in grab the important content instead of reading complete document that would save the time and effort of human. The application of text summarization is helpful in many scenarios like, article summarization, book summarization, research paper summarization, medical document summarization, government document summarizations. In this work we emphasis on the review of text summarization and its types, challenges of text summarization. Organization of paper is as follows section II represents type and features of text summarization section III represents literature review on text summarization section IV represents the architecture of Automatic Text Summarization V presents the challenges of text summarization.

II. TYPES OF TEXT SUMMARIZATION (TS) AND IT'S FEATURES

Text summarizations are mainly classified into two types, which are abstractive text summarization and extractive text summarization; types of TS are explained below:

A. Abstractive Text Summarization:

Abstractive text summarization generates the summary like human being does, which consist of new phrases and words, it modifies the actual text, for generating abstractive text

summarization generation and compression techniques are required [3].

B. Extractive Text Summarization:

Extractive text summarization extracts the important sentences from the text document and keep them in actual summary without modifying the actual text, In Extractive text summarization sentences are extracted depending on the no of feature consist in that, then sentence scoring, and ranking are done to generate summary [4].

C. Features of TS: Features are important attributes are in text documents. Following are some important features which are considered during making the summary.

a) Positional Feature: most of the time first and last sentence could be containing more essential information, researcher should consider it while making the summary

b) Length: Most of the time length as feature is also matters, short and long length of sentences does not contain important information.

c) Title: Title can be important because it contains the theme of documents.

d) Number Data: numerical data containing sentences always considered in summary because it is crucial to the documents.

e) Phrasal Information: noun, verb and adjective are crucial, it always included in summary generation

III. LITERATURE REVIEW

Following table 1 shows literature review on abstractive and extractive text summarization.

Table 1: shows Literature review

AUTHOR AND YEAR	TECHNIQUES	LACUNA
Soma Shrenikaa et.al.,(2021)[5]	Long Short Term Memory (LSTM), Gated Recurrent Unit	Author needs to differentiate result among Long Short

	(GRU) and Transformers	Term Memory (LSTM), Gated Recurrent Unit (GRU) and Transformers algorithms.
Yuuki Iwasaki et.al.,(2020)[6]	Neural network, BERT, Live door news corpus	It is needed to work on a qualitative evaluation of our model to minimize the grammatical mistakes.
A.V. Patil et.al., (2021) [7]	seq2seq model, LSTM Model	It is needed to work on Beam Search in Decoder to get better summary result.
Ramesh Nallapati et.al., (2016) [8]	Attentional Encoder-Decoder Recurrent Neural Networks,	Author needs to focus on generating summary using multiple sentences.
Hengli Song et.al., (2019) [9]	CNN and DailyMail datasets, LSTM model	It is needed to apply different technique to measure accuracy of result
Rupal Bhargavaa et.al.,(2019)[10]	Recurrent Neural Network	It is needed to test result with large size of dataset
Shirwandkar, N. S et.al.,(2018)[11]	Restricted Boltzmann Machine and Fuzzy Logic	Author needs to focus on multi documents summarization.
Joel Larocca Neto et.al.,(2002)[12]	Naive Bayes algorithm and the C4.5 decision tree algorithm	It is needed to extend the work of text summarization by using different algorithm
Nadeen M Abdelaleem et.al.,(2022)[13]	Neural Networks	It is important to work on generating accurate summary by considering important sentences.
T. Sri Rama Raju et.al.,(2017)[14]	Extractive text summarization	It is important to work on abstractive text summarization for producing summary

IV. GENERALIZED MODEL FOR AUTOMATIC TEXT SUMMARIZATION

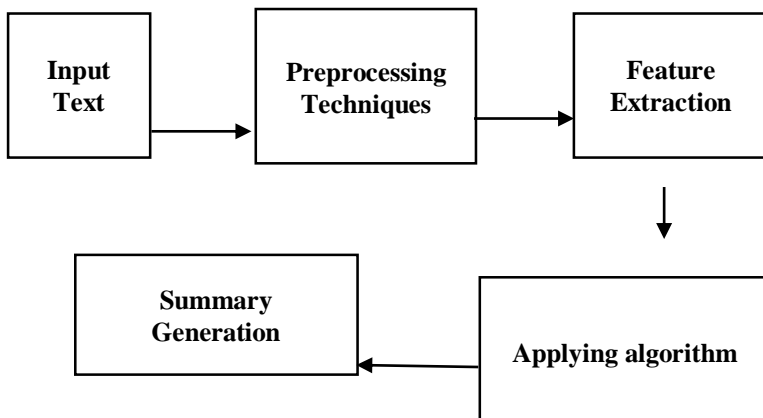


Figure 1: Architecture of Automatic Text Summarization (ATS)

All the phases of ATS are explained below [15].

1)Input Text: In this phase text document is considered

2)Preprocessing Techniques: In this phase text documents preprocessing techniques are applied like tokenization, POS tagging, stop word removal, stemming of word etc.

3)Feature Extraction: Feature extraction is done in this phase by selecting important features, depending upon important features sentences are selected.

4)Applying algorithm: This the crucial phase in which we decide which techniques/algorithms we can apply like extractive text summarization or abstractive text summarization. Algorithms/techniques are applied to select important sentences depending on the no of features.

5)Summary Generation: This phase produces the summary by considering the important sentence only half of the sentences are considered for inclusion in summary.

V. CHALLENGES OF TEXT SUMMARIZATION

Most of the research work is done ATS but some challenges are faced by ATS while summarization, followings are significant challenges are explained below.

1) Evaluation issues: ROUGE, BLEU are automatic evaluation techniques mainly used with extractive text summarization produces poor result as compared to human evaluation score.

2) Issues in important sentences selection: There is difficult task for selecting important sentences to generate summary, there is need to have some standard benchmarks in consideration of summary

3) Lack of training data in different scenarios: there are different dataset are available like cricket, politics etc. same techniques of text summarization cannot cover different method for training dataset.

4) Inclusion of long sentences: Most of the researcher works only on short sentences to produce the summary, because algorithm did not work properly to generate summary by considering the long sentences. It is important to work on such techniques/algorithm which generates summary of long sentences.

5) Need of predefined template: large amount of work is done using ATS, but ATS algorithms did not produces new sentences as their own, there is need to have the predefined template for specific summarization work.

VI. CONCLUSION

Automatic text summarization is important task due to its need, this is the old topic, but still getting interest from researchers, this paper covers review on extractive and abstractive text summarization, this survey shows that extractive text summarization is easier as compared to abstractive text summarization. This paper also presents challenges in Automatic text summarization which will be the helpful for new researcher to solve challenges in ATS domain, in future we are going to work on by combining

abstractive and extractive method for document summarization.

REFERENCES

- [1] Sakshi Bhalla, Roma Verma, Kusum Madaan , “*Comparative Analysis of Text Summarisation Techniques*”, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181,2017.
- [2] Arpita Sahoo, Dr.Ajit Kumar Nayak,“*Review Paper on Extractive Text Summarization*”, International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 5, Issue 4, April 2018
- [3] Ujjwal Rani, Karambir Bidhan,“*Review Paper on Automatic Text Summarization*”, International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 04 | Apr 2020
- [4] R. Ferreira, L. De Souza Cabral, R. D. Lins, G. Pereira E Silva, F. Freitas, G. D. C. Cavalcanti, R. Lima, S. J. Simske, and L. Favaro, “*Assessing sentence scoring techniques for extractive text summarization,*” Expert Systems with Applications, vol. 40, no. 14, pp. 5755–5764, 2013.
- [5] Soma Shrenikaa, Gouri Priya Raminia, Lunsavath Bhagavathi Devia, B Geeta vanib, “*Abstractive Text Summarization By Using Deep Learning Models*”, Turkish Journal of Computer and Mathematics Education Vol.12 No.14 (2021), 2404– 2411.
- [6] Yuuki Iwasaki , Akihiro Yamashita¹, Yoko Konno, Katsushi Matsubayashi, “*Japanese Abstractive Text Summarization using BERT*”, Advances in Science, Technology and Engineering Systems Journal Vol. 5, No. 6, 1674-1682, 2020.
- [7] A.V. Patil, Vikas Raskar, Prachi Pophale, Pooja Fattepure, Umera Attar “*Abstractive Text Summarization Using Attention Mechanism*”, International Journal of Creative Research Thoughts, Volume 9, Issue 5 May 2021 | ISSN: 2320-2882
- [8] Ramesh Nallapati, Bowen Zhou, Cicero Santos, “*Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond*”, SIGNLL Conference on Computational Natural Language Learning (CoNLL), 2016.
- [9] Hengli Song, Haitao Huang, Tongxiao Ruan ,“*Abstractive text summarization using LSTM-CNN based deep learning*”, Multimedia Tools and Applications volume 78, pages857–875 ,2019.
- [10] Rupal Bhargavaa, Yashvardhan Sharmaa ,“*Deep Extractive Text Summarization*”, International Conference on Computational Intelligence and Data Science ,ICCIDS 2019.
- [11] Shirwandkar, N. S., & Kulkarni, S., “*Extractive Text Summarization Using Deep Learning*”, Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018.
- [12] Joel Larocca Neto, Alex A. Freitas, Celso A. A. Kaestner “*Automatic Text Summarization using a Machine Learning Approach*”, Advances in Artificial Intelligence, 16th Brazilian Symposium on Artificial Intelligence, SBIA 2002, Porto de Galinhas/Recife, Brazil, November 11-14, 2002.
- [13] Nadeen M Abdelaleem , Hatem M Abdal Kader , Rashed K Salem and Diaa Salama Abd Elminaam, “*Extractive Text Summarization Using Neural Network*”, Sustainable Economic Development and Advancing Education Excellence in the Era of Global Pandemic,24 Feb. 2022
- [14] T. Sri Rama Raju, Bhargav Allarpu, “*Text Summarization using Sentence Scoring Method*”, International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395 -0056, Volume: 04 Issue: 04 | Apr -2017
- [15] M. F. Mridha , Aklima Akter Lima , Kamruddin Nur , Sujoy Chandra Das , Mahmud Hasan, And Muhammad Mohsin Kabir, “*A Survey of Automatic Text Summarization: Progress, Process and Challenges*”,IEEE Access,22 Nov 2021.