Text Summarization Techniques Algorithm

Dr.NATHALIE JOHN
R. C. Patel IMRD, Shirpur

Abstract: -

In this paper researcher study and review various research on the text summarization. The researcher also discovers algorithm to display text summary in proper format. Text summarization is the problem of creating a short, accurate, and fluent summary of a longer text document. This review determines how to extract proper, short, accurate, and fluent summary reports using current algorithm. In this review paper to study about how to get a summary when we input any long document or multiple documents using algorithm.

Keywords: Natural Language Processing (NLP), Text Summarization Techniques, Extractive summarization, Abstractive summarization and Machine Learning algorithms.

1) Introduction

In this paper researcher studied how to extract needful information from long documents. Text summarization is a process of extracting or collecting important information from original text and presenting that information in the form of a summary. Text summarization has become the necessity of many applications for example search engine, business analysis, market review. Summarization helps to gain required information in less time. NLP tasks are machine translation, tokenization sentence segmentation, named entity recognition, parsing, Part-of-speech tagging and so on. The applications of NLP are knowledge management, language translation, information retrieval, question & answering, automatic summarization etc.

Text summarization approaches can be broadly divided into eight groups: Deep learning techniques, Machine Learning, Graph based, Rule based, Heuristic based, Linguistic Features, Sentence Extraction Method and Sentence weight. These techniques are useful for extracting proper text summary reports.

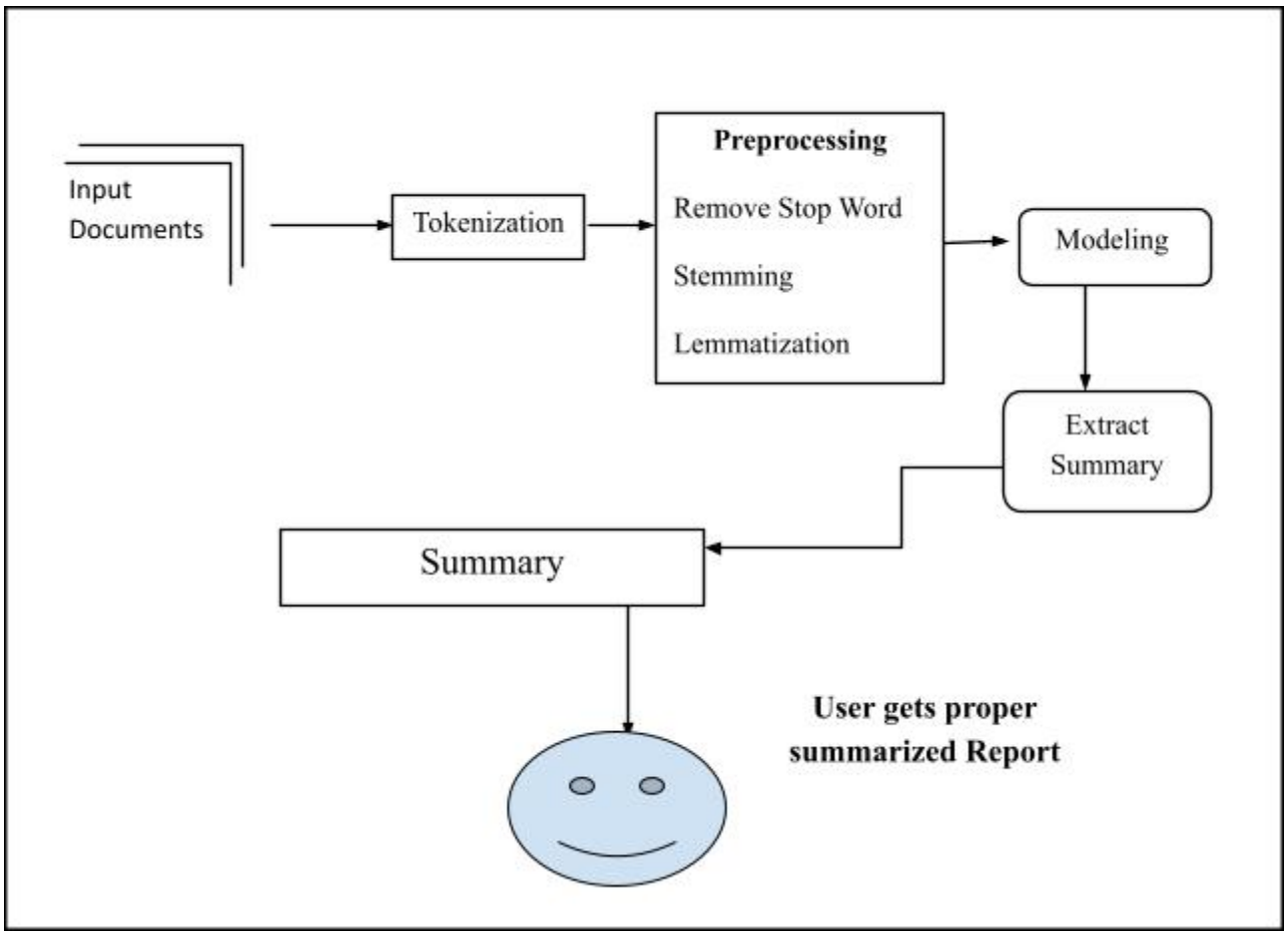


Fig. 1 Text Summarization General Process

In Fig.1 researcher proposes preprocessing and approaches for text summarization of long documents. An extract summary and users get proper summarized reports.

2) Literature Review

Researcher had done literature review because it is most important to all research work to gain the proper result of study.

Researcher doing the work for following purpose:

-To identify gaps in the literature

- To identify the opinions of various researchers.

- To increase knowledge about the subject area.

Sonali Agrawal et.al(2020) Proposed approach deals with text summarization based on a deep learning approach which includes the RBM algorithm for getting better efficiency.

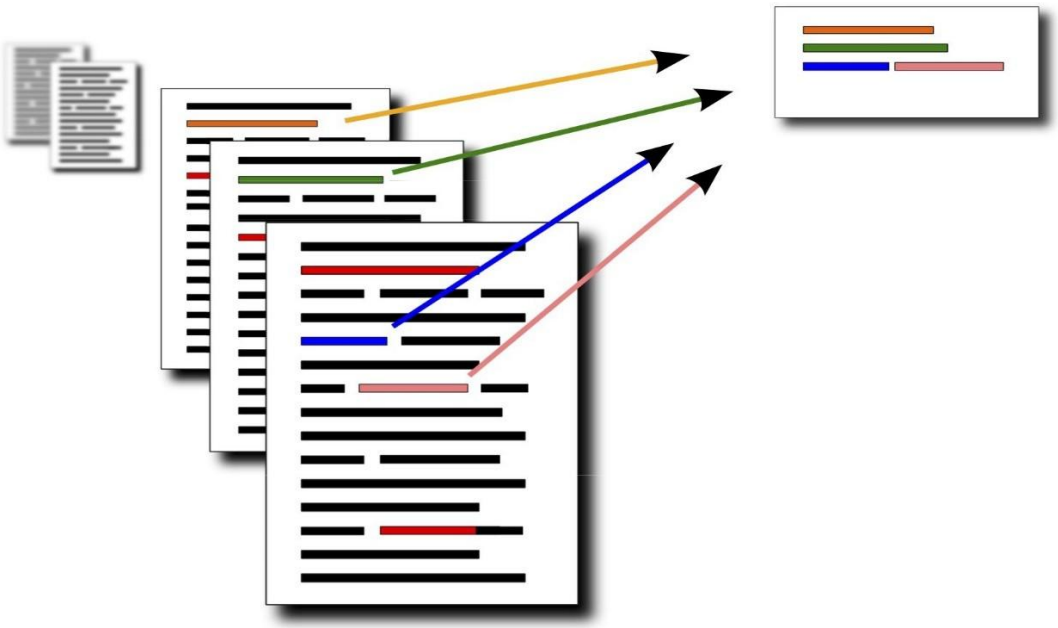
Aakash Sinha et.al(2018), Proposed a fully data-driven approach using feed forward neural networks for single document summarization

Saketh Mattupalli et.al(2020),Discussed various types of text summarization using Pointer Generator network, Text Rank algorithm , seq2seq model and built a model which can perform Abstractive Text summarization.seq2seq model produced better results than the other models

Amita Arora et.al(2017),Presented the results of our research on extractive summarization with a method based on Support Vector Machines. The SVMs were trained using the DUC-2002 dataset and the importance of sentences is judged based on salient features. ROUGE scores are used to compare the system generated summaries with the human generated summaries, and the experimental results show that the system's performance achieved high metrics.

3) Algorithm

In this paper to say how the text summarization is done. Here we developed one algorithm using that algorithm we to manage proper summary documentation using the algorithm.



Input

A=Whole File / Paragraph

B is array;

C=Stack of paraphrase Words.

Algo.

1. B={to split paragraphs and sentences into smaller units.} from A
2. Perform stemming and lemmatization preprocessing on B
3. loop B[i] to B.length
    a. loop 1 to c.length
    b. if B[i]==C then
    c. x=find paraphrase words.
    d. else:
    e. d=find synonym of B;
        i. d==B then
        ii. x=find paraphrase words.

    f. End Loop
4. Recompute and Store x into stack
5. i=i+1;
6. End loop
7. Print x;
8. END

4) Experimental Work

By extensive execution of code and analysis some of the important techniques for Text Summarization from long documents. It is concluded that each line of code has some positive impact for find proper text summarization. Text summarization as discussed is broadly divided into Deep Learning, Machine Learning, Graph Based, Rule Based, Linguistic Features, Heuristic-Based, Sentence Extraction Method and Sentence weight.

The algorithm code design in php. 100 pages data pass through the code and finally produce one proper document that has correct with all the paraphrasing words.

5) FINDINGS: -

After executing an algorithm to find that, regarding the impact of Text summarization techniques is positive to extract the proper summary reports. The algorithm produces proper output to the user as per requirements of summary data.

5) Conclusion:

This paper has attempted for the purpose of Text summarization techniques to create a short, accurate, and fluent summary of a longer text document. The proposed algorithm was successfully tested. The results which were obtained after the analysis were satisfactory and contained valuable information about Text summarization. Thus, text summarization is necessitated and utilized by business analysts, marketing executive, development, researchers, government organizations, students and teachers also. So, this paper gives a summarization of the different NLP and Machine learning techniques with their functionality for Text summarization.

References: -

1) Sonali Agarwal, Pranita Redkar," Text Summarization", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181
2) Akanksha Joshi, E Fidalgo, E Alegre, U de León," Deep learning based text summarization: approaches, databases and evaluation measures",2018 Proceedings of the International Conference of Applications of Intelligent Systems.
3) Aakash Sinha, Abhishek Yadav, Akshay Gahlot," Extractive Text Summarization using Neural Networks", eprint arXiv:1802.10137,2018
4) Saketh Mattupalli, Apurva Bhandari, B.J Praveena," Text Summarization using Deep Learning", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-9, Issue-1, May 2020
5) Amita Arora ,Akanksha Diwedy ,Manjeet Singh and Naresh Chauhan, "Machine Learning Approach for Text Summarization ",International Journal of Database Theory and Application Vol.10, No.8 (2017), pp.83-90

6) Khulood Abu Maria, Khalid Mohammad Jaber and Mossab Nabil Ibrahim," A NEW MODEL FOR ARABIC MULTI-DOCUMENT TEXT SUMMARIZATION", International Journal of Innovative Computing, Information and Control, Volume 14, Number 4, August 2018
7) Virat V. Giri, Dr.M.M. Math and Dr. U. P. Kulkarni , A Survey of Automatic Text Summarization System for Different Regional Languages in India Bonfring International Journal of Software Engineering and Soft Computing, Vol. 6, Special Issue, October 2016
8) Athira S*, Sruthy Manmadhan**, "Deep Learning in Text Summarization - A Survey", Alliance International Conference on Artificial Intelligence and Machine Learning (AICAAM), April 2019
9) S.Sudha Lakshmi 1, Dr.M.Usha Rani, "Multi-Document Text Summarization Using Deep Learning Algorithm With Fuzzy Logic", 2018 IADS International Conference on Computing, Communications & Data Engineering (CCODE).
10) Saketh Mattupalli, Apurva Bhandari, B.J Praveena, "Text Summarization using Deep Learning", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-9, Issue-1, May 2020,
11) Mythreagi. R1*, Dr.N.Yuvaraj2, "Automatic Document Summarization Using Deep Learning Mechanism with Competent Analysis" International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 7 (2019)
12) Geetha J K, Deepamala N, "KANNADA TEXT SUMMARIZATION USING LATENT SEMANTIC ANALYSIS ",2015 International Conference on Advances in Computing, Communications and Informatics (ICACCl), 978-1-4799-8792-4/15
13) Sunitha.C, Dr.A.Jaya, Amal Ganesh, "A Study on Abstractive Summarization Techniques in Indian Languages" Fourth International Conference on Recent Trends in Computer Science & Engineering, Chennai, Tamil Nadu, India, Procedia Computer Science 87 (2016) 25 - 31
14) Dani Gunawan, Siti Hazizah Harahap, Romi Fadillah Rahmat, "Multi-document Summarization by using Text Rank and Maximal Marginal Relevance for Text in Bahasa Indonesia", 2019 International Conference on ICT for Smart Society (ICISS). 978-1-7281-4880-9/19
15) Yuuki Iwasaki, Akihiro Yamashita, "Japanese abstractive text summarization using BERT",*2019 International Conference on Technologies and Applications of Artificial Intelligence (TAAI).* 978-1-7281-4666-9/19
16) Sheikh Abujar, Abu Kaisar Mohammad Masum, Md Mohibullah, Ohidujjaman, "An Approach for Bengali Text Summarization using Word2Vector", 10th ICCCNT 2019 July 6-8, 2019, IIT - Kanpur, Kanpur, India