INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

Text Summarization using Natural Language Processing

M. Raja Kumari [1], Parveen Pathan [2]
[1] Assistant professor, Assistant Professor [2]
Department of Computer Science Engineering
RISE Krishna Sai Prakasam Group of Institutions

Abstract- In this project, Automatic text summarization is basically summarizing of the given paragraph using natural language processing and machine learning. There has been an explosion in the amount of text data from a variety of sources. This volume of text is an invaluable source of information and knowledge which needs to be effectively summarized to be useful. In this review, the main approaches to automatic text summarization are described. We review the different processes for summarization and describe the effectiveness and shortcomings of the different methods. The system works by assigning scores to sentences in the document to be summarized, and using the highest scoring sentences in the summary. Score values are based on features extracted from the sentence. A linear combination of feature scores is used. Almost all of the mappings from feature to score and the coefficient values in the linear combination are derived from a training corpus. Some anaphor resolution is performed. The system was submitted to the Document Understanding Conference for evaluation. In addition to basic summarization, some attempt is made to address the issue of targeting the text at the user. The intended user is considered to have little background knowledge or reading ability. The system helps by simplifying the individual words used in the summary and by drawing the pre-requisite background information from the web.

1. Introduction

In the modern Internet age, textual data is ever increasing. Need some way to condense this data while preserving the information and meaning. We need to summarize textual data for that. Text summarization is the process of automatically generating natural language summaries from an input document while retaining the important points. It would help in easy and fast retrieval of information. There are two prominent types of summarization algorithms. • Extractive summarization systems form summaries by copying parts of the source text through some measure of importance and then combine those part/sentences together to render a summary. Importance of sentence is based on linguistic and statistical features. • Abstractive

ISSN: 2057-5688

summarization systems generate new phrases, possibly rephrasing or using words that were not in the original text. Naturally abstractive approaches are harder. For perfect abstractive summary, the model has to first truly understand the document and then try to express that understanding in short possibly using new words and phrases. Much harder than extractive. Has complex capabilities like generalization, paraphrasing and incorporating realworld knowledge. Majority of the work has traditionally focused on extractive approaches due to the easy of defining hard-coded rules to select important sentences than generate new ones. Also, it promises grammatically correct and coherent summary. But they often don't summarize long and complex texts well as they are very restrictive.

2. Potential applications

The objective of the project is to understand the concepts of natural language processing and creating a tool for text summarization. The concern in automatic summarization is increasing broadly so the manual work is removed. The project concentrates creating a tool which automatically summarizes the document.

Scope

The project is wide in scope | all of the limitations stated below may seem to contradict that, but they are the only restrictions applied. This project looks at single document summarization - the area of multi document summarization is not covered. Also, the summaries produced are largely extracts of the document being summarized, rather than newly generated abstracts. The parameters used are optimal for news articles, although that can be changed easily.

3. Methodologies

For obtaining automatic text summarization, there are basically two major techniques i.e.- Abstraction based Text Summarization and Extraction based Text Summarization. Extraction Based Extraction The Extractive summaries are used to highlight the words which are relevant, from input source document. Summaries help in generating concatenated sentences taken as per the appearance. Decision is made based on every sentence if that particular sentence will be included in the summary or not. For example, Search engines typically use Extractive summary generation methods to generate summaries from web page. Many types of logical and mathematical formulations have been used to create summary. The regions are scored and the words containing highest score are taken into the consideration. In extraction only important sentences are selected. This approach is easier to implement. There are three main obstacles for extractive approach. The first thing is ranking problem which includes ranking of the word. The second one selection problem that includes the selection of subset of particular units of

ISSN: 2057-5688

ranks and the third one is coherence that is to know to select various units from understandable summary. There are many algorithms which are used to solve ranking problem. The two obstacles i.e. - selection and coherence are further solved to improve diversity and helps in minimizing the redundancy and pickup the lines which are important. Each sentence is scored and arranged in decreasing order according to the score. It is not trivial problem which helps in selecting the subsets of sentences for coherent summary. It helps in reduction of redundancy. When the list is put in ordered manner than the first sentence is the most important sentence which helps in forming the summary. The sentence having the highest similarity is selected in next step is picked from the top half of the list. The process has to be repeated until the limit is reached and relevant summary is generated. People by and large utilize abstractive outlines. In the wake of perusing content, Individuals comprehend the point and compose a short outline in their own particular manner creating their very own sentences without losing any essential data. In any case, it is troublesome for machine to make abstractive synopses. Along these lines, it very well may be said that the objective of reflection based outline is to make a synopsis utilizing regular dialect preparing procedure which is utilized to make new sentences that are syntactically right. Abstractive rundown age is difficult than extractive technique as it needs a semantic comprehension of the content to be encouraged into the Common Dialect framework. Sentence Combination being the significant issue here offers ascend to irregularity in the produced outline, as it's anything but an all around created field yet. Abstractive arrangement to grouping models is by and large prepared on titles and captions. The comparative methodology is embraced with archive setting which helps in scaling. Further every one of the sentences is revamped in the request amid the inference. Document synopsis can be changed over to regulated or semi-administered learning issue. In directed learning methodologies, indications or signs, for example, key-phrases, point words, boycott words, are utilized to recognize the sentences as positive or negative classes or the sentences are physically labelled. At that point the parallel more tasteful can be prepared for getting the scores or synopsis of each sentence. Anyway they are not effective in removing archive explicit summaries. If the report level data isn't given then these methodologies give same expectation independent of the record. Giving archive setting in the models diminishes this issue.

4. LITURATURE SURVEY

• “An approach to sentence-selection-based text summarization”, Fang Chen; Kesong Han; Guilin Chen is a author of this paper, this paper published in 2016. This paper presented an We introduced a newly

ISSN: 2057-5688

developed text summarization system. It supports both Chinese and English, while this paper focuses on Chinese processing. We apply 6 word level features and 3 sentence level features to weigh each word and sentence. We also describe two new techniques, one is for processing the topic sensitive word feature and another is for processing the sentence length feature. Primary subjective evaluation shows that these approaches are effective and efficient, and performance of the system is promising.

• "Automatic Text Summarization Using Hybrid Fuzzy GA-GP" is paper of A. KianiB; M.R. Akbarzadeh-T , 2015 A novel technique is proposed for summarizing text using a combination of Genetic Algorithms (GA) and Genetic Programming (GP) to optimize rule sets and membership functions of fuzzy systems. The novelty of the proposed algorithm is that fuzzy system is optimized for extractive based text 3 Metadata extraction from scientific pdf summarizing. In this method GP is used for structural part and GA for the string part (Membership functions). The goal is to develop an optimal intelligent system to extract important sentences in the texts by reducing the redundancy of data. The method is applied in 3 test documents and compared with the standard fuzzy systems as well as two other commercial summarizers: Microsoft word and Copernic Summarizer. Simulations demonstrate several significant improvements with the proposed approach.

• "Generic text summarization using local and global properties of sentences" C. Kruengkrai; C. Jaruskulchai; 2015. In this paper described The paper With the proliferation of text data on the World-Wide Web, the development of methods for automatically summarizing these data becomes more important. Here, we propose a practical approach for extracting the most relevant sentences from the original document to form a summary. The idea of our approach is to exploit both the local and global properties of sentences. The local property can be considered as clusters of significant words within each sentence, while the global property can be thought of as relations of all sentences in the document. These two properties are combined to get a single measure reflecting the in formativeness of sentences. Experimental results show that our approach compares favorably to a commercial text summarizer.

• "A Review on Optical Character Recognition and Text to Speech Conversion" Swati Vikas Kodgire; 2013. The application depending on image and voice with a parallel functioning is suitable to assist physically challenged people. So that dependability of a challenged person is decreased to a improved level. Image acquisition based text reader can help visually challenged people to manage the handheld objects in day to day life. Initially steps involves capturing of image, distinguishing image with text portion and residual regions, image pre-processing on

INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

region of interest, after the extraction of characters and words, conversion of text to speech is done. To splinter text from a document it is obligatory to discover all the possible manuscript text regions. Text detection, line detection, character identification, feature extraction, training of extracted features are the steps in sequence that are executed.

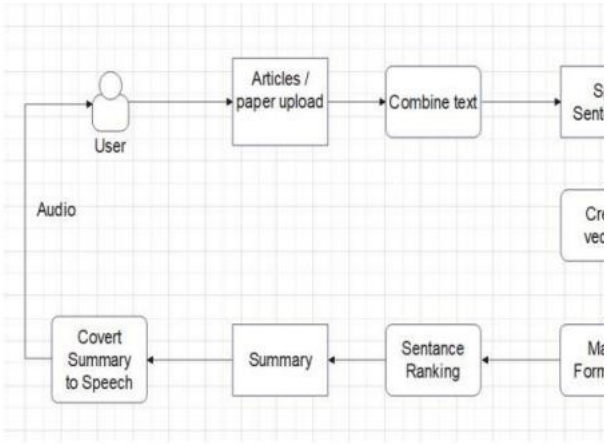
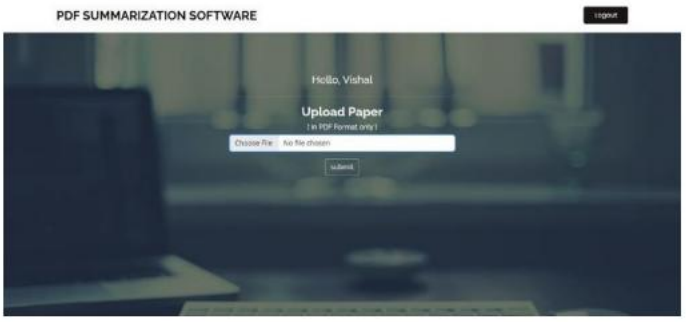


Fig 1: SYSTEM ARCHITECTURE

5. RESULT



6. CONCLUSION

ISSN: 2057-5688

The Conclusion of this project is that the client will get a web application that will execute on client side and get the summary of the input document as per clients requirement. The automatic generated summary is useful for the client to understand the core concept of the document with in few lines instead of reading whole document. It was seen that this code performs really well in reading straightforward PDF text files. Should enable users to select the desired PDF and convert it to audio and display text in, so the user can understand that particular text has been read. Should enable students with reading disabilities. The success of this research project is significant given the broad use of audiobooks in literacy and library programs across the United States. Teachers and school librarians may also use these findings as a rationale for adding audiobooks to the list of reading strategies used successfully with struggling readers

REFERENCES:

[1]. Pankaj Gupta, Vijay Shankar Pendhluri, Ishant Vats, "Summarizing text by ranking text units according to shallow linguistic features", Feb. 13 16, 2011 ICACT, 2011.

[2]. Rajesh S. Prasad, U. V. Kulkarni, Jayashree R. Prasad, "Connectionist Approach to Generic Text Summarization,", World Academy of Science, Engineering and Technology 55,2009.

[3]. R. S. Prasad, U. V. Kulkarni, J. R. Prasad, "A Novel Evolutionary Connectionist Text Summarizer (ECTS),", 2009,IEEE Xplore.

[4]. Rajesh Shardanand Prasad, Uday. V. Kulkarni, "Implementation and Evaluation of Evolutionary Connectionist Approaches to Automated Text Summarization", Journal of Computer Science 6 (11): 1366-1376, 2010 ISSN 1549-3636, 2010 Science Publications.

[5]. Ranjit Bose "Natural Language Processing: Current state and future directions", International Journal of the Computer, the Internet and Management Vol. 121, January – April, 2004.

[6]. Natural Language Processing Techniques Applied in Information Retrieval-Analysis and Implementation in Python, TulikaNarang, International Journal of Innovations Advancement in Computer Science IJIACS ISSN 2347 – 8616 Volume 5, Issue 4 April 2016.

[7]. Pdf. (2021, March 08). Retrieved March 09, 2021, from https://en.wikipedia.org/wiki/PDF

[8]. 7 ways Audio books benefit students who struggle with reading. (n.d.). Retrieved March 09, 2021, from: :https://learningally.org/Solutions-for School/7-Ways-Audio books-Benefit-Students WhoStruggleWith-Reading.

ISSN: 2057-5688

[9]. Dr. B Sankara Babu, Srikanth Bethu, K. Saikumar, G. Jagga Rao, "Multispectral Satellite Image Compression Using Random Forest Optimization Techniques" Journal of Xidian University, in Volume 14, Issue 5-2020.

[10]. G. Jagga Rao, Y. Chalapathi Rao, "Human Body Parts Extraction in Images Using JAG-Human Body Detection (JAG-HBD) Algorithm Through MATLAB" Alochana Chakra Journal, Volume IX, Issue V, May/2020.

[11]. Dr. k. Raju, A. Sampath Dakshina Murthy, Dr. B. Chinna Rao,G. Jagga Rao "A Robust and Accurate Video Watermarking System Based On SVD Hybridation For Performance Assessment" International Journal of Engineering Trends and Technology (IJETT) – Volume 68 Issue 7 - July 2020.

[12]. G. Jagga Rao, Y. Chalapathi Rao, Dr. Anupama Desh Pande "A Study of Future Wireless Communication: 6G Technology Era " volume 14, issue 11,2020.

[13]. G. Jagga Rao, Y. Chalapathi Rao, Dr. Anupama Desh Pande "Deep Learning and AI-Based millimeter Wave Beamforming Selection for 6G With Sub-6 GHz Channel Information" Volume 21 : Issue 11 – 2020.

[14]. G. Jagga Rao, Y. Chalapathi Rao, Gopathi Shobha,G. Jagga Rao,P. Lavanya, M. Ravi “Deep Learning based Millimeter Wave/Sub - THz RMIMO-OFDM Systems with Beamforming Wireless communication ” International Journal of Membrane Science and Technology, 2023, Vol. 10, - 2023.