NATURAL LANGUAGE PROCESSING USING FUZZY PROCESSING FOR ENGLISH TEXT SUMMARIZATION

Ms. Seema Sachin Vanjire[1], R. B. Dhumale[2], Mrs. M. V. Shelke[3], Dr. Amol A. Bhosle[4], Dr. Ashutosh M. Kulkarni[5]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Nowadays, Text summarization is significant application for documents, research papers and online news. The objective of this paper is to decrease document size and extracting the most important part of it. Summarization helps to gain required information without going through whole document. This proposed work is to summarize and present summary of text using an extractive approach with fuzzy logic. Overall, an extractive approach of summarization is the subset of content which sometimes will not assure satisfactory output. So along with this fuzzy logic helps to improve in getting outputs. In fuzzy summarization process followed by: Pre-processing (segmentation of sentences, tokenization, get rid of stop words), Feature-collection, Sentence Score, Sentence Rank and Summary collection**.**

Keywords: Fuzzy Processing, Frequency, Natural Language processing, Summarization, Tokenization

[1]Department of Computer Engineering Vishwakarma University, Pune, India

[2,3]AISSMS Institute of Information Technology, Pune, India

[4]School of Computer EngineeringMIT Academy of Engineering, Alandi Pune, India

[5]Vishwakarma Institute of Technology, Pune, India

Email: [1]seema.vanjire@vupune.ac.in, [2]rbd.scoe@gmail.com, [3]mayura.shelke@aissmsioit.org. [4]amolabhosle@gmail.com, [5]ashutosh.kulkarni@vit.edu

DOI: 10.31838/ecb/2023.12.s3.163

1. Introduction

Text Summarization is extracting necessary data from a large amount of information. The amount of available data is increasing every day, so managing that data is a kind of problem in different sectors. Basically, Summary has less size than original document which contains only important sentences of data. Most of the time it is difficult to generate summary of every document manually which is available. For this, the alternative can be like generate automatic process for summary construction.

The two main types of Automatic text summarization: extraction and abstraction. [A]The extractive method gives us the significant phrases and lines having the highest score from the primary text according to feature score. It then combines all the top n number lines to give the summary. So, in this case, every line and word of the summary is a subset of the original document which is summarized. [B] Abstractive text summarization is a slightly advanced method. It creates entirely new phrases and sentences as per human understanding. This is a difficult approach but this approach is likely to be used by humans. In the Extraction method for summarization, an important and fundamental step is identifying the features.

The approach of this proposed work uses the terms frequency and weighted frequency attribute combination technique to choose the score of words and sentences. So, this proposed work uses text summarization depending on fuzzy logic to extricate an abstract of data. Proposed system attempts to summarize data in three different types of input i.e., file, browser links and normal texts.

LITERATURE REVIEW

[1] Rani, Ujjwal, and Karambir Bidhan disclosed a review of different approaches of text summarization like abstractive approach- Structure Based Approach, Semantic Based approaches and Extractive approach -Fuzzy Logic for summarization, TF- IDF method, Query Based Method, Graph Theoretic Method, Cluster Method, Machine Learning, Latent Semantic Analysis Method, Text Summarization using Neural Network.[1]

[2] Neelima, G., et al disclosed method of extractive Text Summarization utilizing Deep Natural Language Fuzzy Processing. Naive- Bayes' approach has been used in order to determine significant sentences of a text. For this approach, steps followed are tokenization, pre-processing, stemming, fuzzy processing [2]

[3] Al Qassem, Lamees, et al. discloses an Automatic Text Summarization using Fuzzy Logic in Arabic language. This is a smart Arabic summarization system that has great accuracy as compared to current. The system focuses to give summaries from to user.[3]

[4] Devihosur, Pratibha, and R. Naseer discloses method that utilizes an unsupervised learning approach. This system is using a solitary or single input content which is going to be outlined with the help of given summarization rate. Also, according to a particular rate of summarization at a specific occurrence, some sentences are taken as an outline [4]

[5] Deepali, K. Gaikwad, and C. Namrata Mahender focuses on comparative review on two approaches of summarization and methods of it. The approaches in this paper are –A] Abstractive summarization: 1. Structured based Approaches (Ontology Method, Tree Based, Template Based, Lead and Body Phrase Method, Rule Based Method) 2. Semantic based approaches (Multi semantic model, Information Item Based Method, Semantic Graph Based Method) Extractive Summarization [5]

[6] Christian, Hans, et. al focuses on the extractive type of summarization approach. This system uses TF-IDF to build summary. This system uses TF-IDF to build a summary. The three different documents are provided to be summarized and to calculate its accuracy. An analysis is performed to find out how the program can reach a certain precision.[6]

[7] Dixit, Rucha S., and S. S. Apte focuses on enhancement of text summarization by making use of fuzzy logic. The proposed method utilizes feature extraction method by using fuzzy rules and sets in order to select sentences according to their features. Fuzzy logic provides decision with strong reasoning.[7]

Following Table I shows and discussed about the different algorithms and techniques proposed by various authors and challenges observed.

TABLE I: REVIEW LITERATURE OF VARIOUS EXISTING TECHNIQUES FOR SUMMARIZATION.

Paper	**Type of summarization**	**Algorithm Used**	**Summary**	**Challenges and Discussions**
[1]	Abstractive and extractive summarization	Semantic Based approaches, Semantic Based, Cluster Method, ML, Latent semantic method	The paper reviews both extractive and abstractive approaches along with procedures used.	The Abstract approach is a bit more complex than the extractive approach.
[2]	Extractive summarization	Fuzzy logic with naïve-Bayes algorithm.	Fuzzy logic based on deep learning helpful in generating accurate summary.	Addition of support of other file formats like .docx, .pdf
[3]	Extractive summarization	Fuzzy Logic method	Fuzzy Logic improved summarization performance of Arabic System.	Looking for better pre-processing for more accuracy of Arabic Summarizer.
[4]	Extractive summarization	Unsupervised learning with Streamlined Lesk Calculation	Summary is generated depending upon the semantic data.	Attempting summarization method with more tongue specific segments.
[5]	Abstractive and extractive summarization	Structured Based Approach, Semantic Based Approach, Extractive approach methods.	Paper reviews both extractive and abstractive approaches along with advantages and disadvantages of each approach.	Very less work and research has been done with an abstractive approach in Indian language.
[6]	Extractive summarization	TF-IDF	The accuracy of the summary generated is 67%.	Try to make a summary based on the title in order to achieve better accuracy.
[7]	Extractive summarization	Fuzzy logic feature-based method	System has been examined with 30 news documents and comparison done with Copernic summarizer and MS Word 2007 summarizer	Extension of use of proposed method with multi-documents.

PROPOSED SYSTEM

A. Proposed Work:

The proposed system has two parts in total i.e., GUI and Processing unit. GUI is for effective presentation of system. In processing unit actual working of text summarization is takes place. It is composed of three stages for Text Summarization.

- Pre-Processing
- Calculating frequency
- Summary Generation

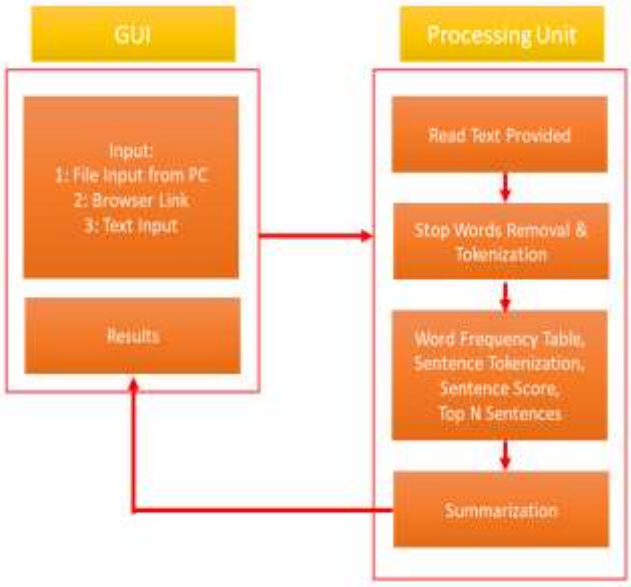

Fig.1. Proposed Methodology of Text Summarization

2. Methodology

a) Removal of stop Words: This step includes removal of stop words as per NLP. Basically, stop words do not contribute to the actual context of the document. For example, in case detection of emotion system, some of the words like "is", "am", and "the" are not important in detecting of emotions as they do not convey toward emotion. In the sentence "I am feeling happy", the words "I" and "am" this can be draw out because they are not related to emotion. It suggestible to have just those words that are necessary in system, so that processing and calculating results can be saved. So that we can have more robust NLP engine.

b) Tokenization: Tokenization means dividing the sentences into chunks of words. This model is used to carry out tasks in NLP. In this proposed work it is used at two levels: first at word-level and second at sentence-level. Word-level tokenization gives a set of phrases in a sentence. Example: I can do it properly. => [‘I’ can’ ‘do’ ‘it’ ‘properly]. In sentence-level it gives a group of sentences from the given input.

c) Fuzzy Processing and Extraction of important sentences: In order to extract important sentences from document there must be some calculation to do. So, for this, the following steps are performed:

Step 1: Count the frequency i.e., occurrence of words from the pre-processed data of a given document.

Step 2: Count the weight frequency for all words. For this step divide each word prevalence with the word having the highest frequency.

Step 3: Tokenize every sentence from the given input text.

Step 4: Count the score of sentences. For the same add the weight frequency of every word in the particular sentence.

Step 5: Arrange the tokenized listing of sentences as in step with their ratings in lowering order.

Step 6: Take out top ‘n’ sentences among the listing which is formed at the above step.

3. Results And Discussion

The finalized proposed system consists of GUI with different pages according to the type of input taken by user.Following Fig 2 shows the first page of GUI which is by default display first at the start of system. On the top of screen, there are various options to give input for system.

Fig 2. Introductory Page

Following Fig 3. shows the GUI page by which input can be given in the form of the document files from pc / laptop by clicking the open file button.

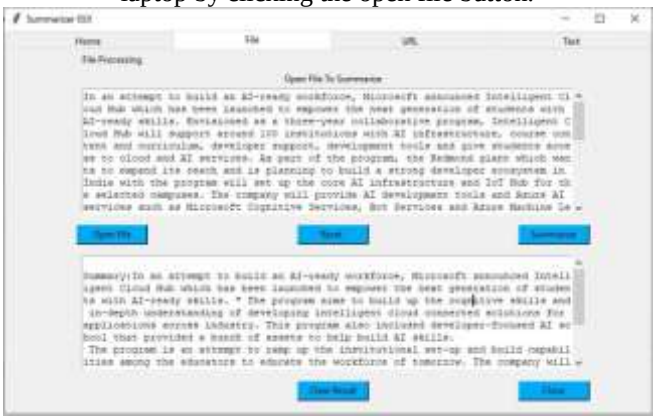

Fig 3. Input From File

Following Fig 4. shows the GUI page by which input can be given through the browser. User has to paste a link in the input field to get a summary.

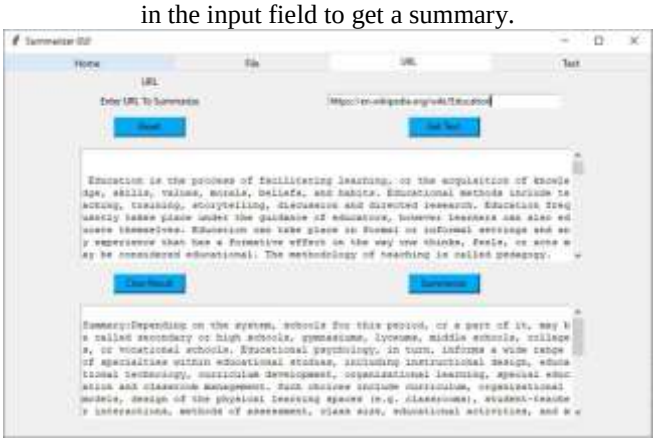

Fig 4. Input From Browser Links

Following Fig 5. shows the GUI page by which text input can be given through the keyboard. User has to paste text in the input field to get summary.

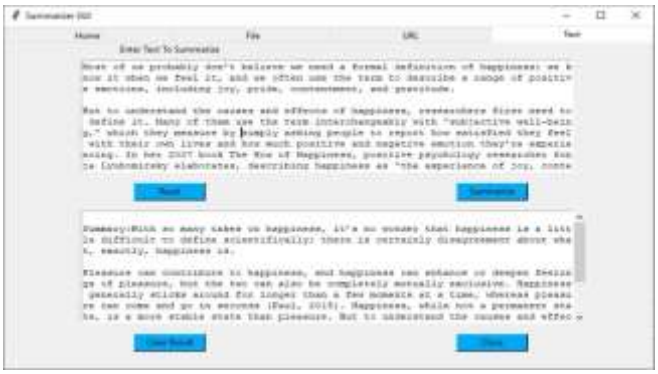

Fig 5. Text Input from Keyboard

4. COMPARISONS

The distribution of techniques for summarization is depends upon the problem with which system has to deal with. Following Table II shows and discussed about the problems and techniques to overcome problems for summarization.[8]

TABLE II: REVIEW OF VARIOUS METHODS AND PROBLEM IN TEXT SUMMARIZATION

Problem	**Technique**	**Methods**
Extraction	Fuzzy-based	Fuzzy Logic
	Rule-based	Rule-based
Semantic	Statistic	Latent semantic analysis (LSA)
	Machine learning	LSA + ANN Deep Learning
Topic modeling	Topic model	Latent Dirichlet Allocation
Similarity	Graphical	Text-rank, lex-rank

From the published literature paper, we conclude that from last ten years the comparison of techniques which has been used for summarization is as follows [8]:

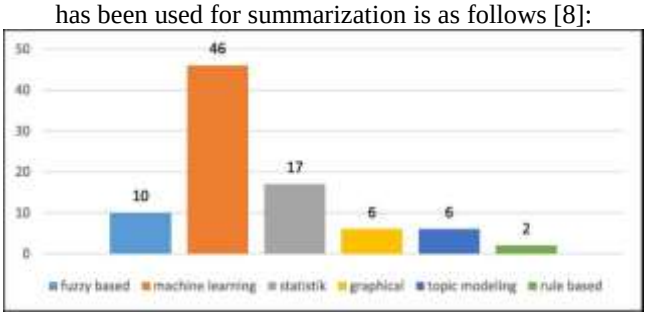

Fig 6. Representation of methods applied in Text Summarization [8]

Limitations

This Proposed system is not able to summarize the data which is from local languages like regional language.

5. Conclusion

"Text summarization" based on natural language processing using fuzzy logic method is useful in summarizing text for English language. The proposed system is successfully able to summarize the data which can be given in the form of text from keyboard, from browser links and in the form of files from pc/laptop. It helps to save time and gathering of information becomes easier with the most important information after summarizing.

Future Scope

The Proposed system is successfully able to summarize data from English Language. The further modifications will include the summarization of local languages by providing required inputs. Also, GUI can be further modified in future according to upcoming changes.

6. References

Rani, Ujjwal, and Karambir Bidhan. "Review paper on automatic text summarization." Int. Res. J. Eng. Technol.(IRJET) 7.04 (2020).

Neelima, G., et al. "Extractive text summarization using deep natural language fuzzy processing." International Journal of Innovative Technology and Exploring Engineering (IJITEE) 8.6S4 (2019).

Al Qassem, Lamees, et al. "Automatic Arabic text summarization based on fuzzy logic." Proceedings of the 3rd international conference on natural language and speech processing. 2019.

Devihosur, Pratibha, and R. Naseer. "Automatic text summarization using natural language processing." International Research Journal of Engineering and Technology (IRJET) 4.08 (2017): 667-73.

Deepali, K. Gaikwad, and C. Namrata Mahender. "A Review Paper on Text Summarization." International Journal of Advanced Research in Computer and Communication Engineering 5.3 (2016).

Christian, Hans, Mikhael Pramodana Agus, and Derwin Suhartono. "Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF)." ComTech: Computer, Mathematics and Engineering Applications 7.4 (2016): 285-294.

Dixit, Rucha S., and S. S. Apte. "Improvement of text summarization using fuzzy logic based method." IOSR Journal of Computer Engineering (IOSRJCE) 5.6 (2012): 5-10.

Suanmali, Ladda, Naomie Salim, and Mohammed Salem Binwahlan. "Feature-based sentence extraction using fuzzy inference rules." 2009 International Conference on Signal Processing Systems. IEEE, 2009.