International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

NEWS TEXT SUMMARIZATION BASED ON MULTI FEATURE AND FUZZY LOGIC

Ms.M.ANITHA[1], Ms. V.RAJA RAJESWARI [2]

#1 Assistant professor in the Master of Computer Applications in the SRK Institute of Technology, Enikepadu, Vijayawada, NTR District

#2 MCA student in the Master of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District

ABSTRACT_ Because the amount of data on the Internet is rapidly increasing, automatic text summarization work has become increasingly important in the last 70 years, and automatic text summarization work can extract useful information and knowledge that users require, which can be easily handled by humans and used for a variety of purposes. News text is the type of text that most people are exposed to, especially in their daily lives. This study introduces a new automatic summarization model for news text that is based on fuzzy logic rules, multi-features, and the Genetic algorithm (GA).

To begin, we score each word and extract words that surpass the preset score as keywords; and because news content is a special sort of text, it has many specific characteristics, such as time, place, and characters, these distinctive news aspects can often be extracted directly as keywords. A genetic algorithm weighs sentence features in the second category. A linear combination of these characteristics demonstrates the significance of each statement.

1.INTRODUCTION

Today's people believe that the most potent social media data of explosive growth is the daily news from Web, WeChat, Weibo, and various types of industry data. In the era of big data, a lot of data is produced on the Internet every day. The volume of data is already much larger than existing storage, processing and analysis tools. Web news has emerged as one of the best media for people to use to stay up to date on current events, which are constantly changing.

Especially for some businesses and people who are in great need of information, people do not have enough time to read all the news online while dealing with these massive amounts of news. Consequently, people's attention has recently turned to research on automatic news summaries. It can both alleviate the issue of information overload on the Internet and make it easier for users to understand the information they receive. Automatic news summaries are produced by taking the most important information from a news article and condensing it using algorithms.

The way people read news online could be completely changed by this technology. By giving users a brief summary of the news, technology can help users save time and avoid information overload. It can also make it easier for readers who have trouble reading lengthy articles or complex language to understand the news. Additionally, this technology enables users to have their news personalised based on their reading preferences and interests, resulting in a more customised news experience. Users may become more engaged as a result, and it may also motivate them to follow current events.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

The so-called "news automatic summary" process involves the computer automatically extracting the summary's relevant content from the original news text and creating a succinct, coherent short article that accurately and completely reflects the news's main points . Some works generate summaries for each individual document on web pages using search engines .

In other works, text analysis techniques are employed to identify the key elements of each document and to label them in order to create a cluster of web documents. Due to their effectiveness in text data mining, Neural Networks and Support Vector Machine techniques have recently been used in text summarization. Additionally, some works use literature summaries, topic detection, and tracking tasks to quickly mine the text's theme. Fuzzy logic has recently become more prevalent in text data mining, and evolutionary algorithms are used in text summarization because they have the ability to deal with complex problems that are challenging for traditional optimization algorithms to solve without being constrained by the nature of the problem.

However, most studies ignore the distinctive characteristics of news articles, namely the three components of time, place, and person, which together make up the fundamental facts of a news story, and instead only take into account the general word features and sentence features when extracting the features of news texts. The majority of studies were based on calculating the importance of the sentences in the text to extract important sentences to form a summary, but these studies ignored the fuzziness of artificial intelligence because different people will have different opinions for the same thing. News as a type of natural language processing (NLP) is simultaneously the embodiment of artificial intelligence, thinking.

As a result, this paper suggests a method for automatic summarization that is based on fuzzy logic and multi-feature extraction. It selects the most significant features through feature selection, chooses the key words by calculating each word's score, and then scores the sentences by combining the unique three components of the news text. Each text feature is given optimised weights by a fuzzy logic system that is then used to score the sentences and extract summaries

2.LITERATURE SURVEY

2.1 J. Leskovec, M. Grobelnik, and N. Milic-Frayling, “Learning substructures of document semantic graphs for document summarization,” in Proceedings of the KDD 2004 Workshop on Link Analysis and Group Detection (LinkKDD), Seattle, WA, USA, 2004.

According to JurijLeskovecet al. , the original document should be represented as a triple of subject-predicate-object sentences in a semantic graph. To produce summaries, the graph's sub-structure is extracted. In order to extract sentences, a set of triplets that contribute to this are identified using an SVM classifier. To improve the performance of the suggested model, a wide range of linguistic attributes are incorporated into it. After being refined through a series of steps involving coreference resolution, cross-sentence pronoun resolution, and sentence normalisation, the set of triplets created in

the previous step are combined to produce a semantic graph. This semantic graph then becomes a useful tool for NLP tasks like sentiment analysis, entity recognition, and topic modeling because it can be used to extract insightful information and knowledge from massive amounts of text data. The model is robust and able to handle a variety of text inputs thanks to the incorporation of various linguistic attributes.The tool also has a user-friendly interface and requires little coding expertise, making it available to a variety of users. In order to meet the needs of a particular project, it also enables customization and fine-tuning.

2.2 N. Munot and P.-N. P. I. Sharvari S. Govilkar Department of Computer Engineering, Mumbai University, "Conceptual framework for abstractive text summarization," International Journal on Natural Language Computing (IJNLC) Vol. 4, No.1, February 2015, 2015.

Summarization was carried out by Nikita Munot et al. so that the documents could be divided into sentences and built using SPO triples to produce a semantic graph known as rich semanticgraph (RSG). After RSG, the graph nodes are reduced to Subject Noun (SN), Main Noun (MN), and Object Noun (ON) (ON). The Semantic Graph can be reduced, and a summary can be generated. Summarization is the process of condensing the semantic graph and producing a summary. Text summarization, news article summarization, and document summarization are just a few examples of the many uses for summarization. It aids in effectively removing the most crucial information from a sizable text corpus. In today's fast-paced world, where people are overloaded with information and need to quickly understand the main points of a document, this technology has grown in importance. By enabling people to quickly locate the most crucial information in a document, summarization can reduce time spent reading and increase productivity. Additionally, summarising a document can aid in decision-making by giving readers a quick overview of its content, which enables them to decide quickly and with more knowledge. Professionals who must daily read and analyse large amounts of text will find this technology especially helpful.

2.3 E. Lloret, E. Boldrini, T. Vodolazova, P. MartÃnez-Barco, R. MuÃ´soz, and M. Palomar, "A novel concept-level approach for ultraconcise opinion summarization," Expert Systems with Applications, vol. 42, no. 20, pp. 7148 – 7156, 2015. [Online].

A model for creating extremely succinct concept-level summaries was created by Lloret.E et al. After performing a lexical analysis, the system first converts the input document into its syntactic representation. The lexical units are used as the input for a language generation tool to generate summaries. This system relies on the presumption that all sentences are anaphora resolved and lacks semantic representation of the text. The generated summaries might not accurately convey the original text's intended meaning, especially when there are several possible readings. As a result, it's crucial to carefully assess the efficacy and accuracy of the summaries that are produced. This demonstrates how crucial human evaluation is to tasks involving natural

language processing, like text summarization. It also highlights the need for more research and development in this area to raise the standard of summaries that are produced. To ensure that the machine-generated summaries adhere to the necessary standards of accuracy and relevance, human evaluation is crucial. Therefore, it is essential to make investments in more cutting-edge technologies that can generate high-quality summaries with little assistance from humans.

3.PROPOSED SYSTEM

To provide precise and useful news summaries, the suggested system combines the advantages of fuzzy logic analysis with multi-feature analysis. By changing the weights given to each score, the system is adaptable to multiple domains and themes and may be used to deliver real-time news summaries for a range of applications. The BART model is employed, and it has been fine-tuned on CNN Daily Mail. It is already trained on the English language.

3.1 IMPLEMENTATION

- Modules and libraries
- Data Preprocessing
- Data Visualization
- Model Building

3.1.1 Modules and libraries:

A library is a group of related functions that can be incorporated into your Python code and used as needed in the same way as other functions. Rewriting code to carry out a common task is not necessary. With libraries, you can effectively increase the functionality of your code by importing pre-existing functions. Using libraries not only reduces time and work requirements, but also guarantees that the code is dependable and has been thoroughly tested by a large developer community, reducing the likelihood of errors. Additionally, rather than wasting time on routine tasks, it enables developers to concentrate on the distinctive aspects of their project.

3.1.2 Data preprocessing:

Pre-processing describes the changes made to our data before we feed it to the algorithm. A technique for turning unclean data into clean data sets is data preprocessing. In other words, data is always gathered from various sources in a raw state that precludes analysis. Data preprocessing entails a number of steps, such as data cleaning, data transformation, and data reduction, to ensure that the data is accurate, consistent, and suitable for analysis. It improves the data's quality and increases the accuracy of the inferences made from it, making it a crucial step in the analysis of data.

3.1.3 Data Visualisation:

Data visualisation is the process of transforming sizable data sets into statistical and graphical representations. In data science and knowledge discovery techniques, it is essential to make data more understandable and accessible. Visual representation is required for charts and graphs in order to facilitate quick information absorption and make them easier to understand. Avoid hesitating on tables with large data sets if you want to keep the audience's interest for a longer period of time

3.1.4 Model Building

Here are the key steps involved in the model building process:

Define the problem: Clearly understand the problem you are trying to solve and

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

determine the type of machine learning task, such as classification, regression, clustering, or recommendation.Collect and preprocess data: Gather relevant data that is representative of the problem domain. Clean the data by handling missing values, removing outliers, and normalizing or scaling the features as necessary. Split the data into training and testing sets.Select a model: Choose an appropriate machine learning algorithm or model based on the problem type, available data, and desired outcome. Some common algorithms include linear regression, decision trees, support vector machines, random forests, and neural networks.

3.2 SYSTEM ARCHITECTURE

The conceptual model, also known as the system architecture, of a system determines its structure, behaviour, and other characteristics. An architecture description is a formal description and representation of a system that is designed to make it easier to analyse its structures and behaviours. System architecture and architecture descriptions are crucial to software engineering because they improve stakeholder communication, spot potential issues, and give a clear understanding of the system's design. More efficient system development, maintenance, and evolution can be achieved with an effective system architecture and architecture description.

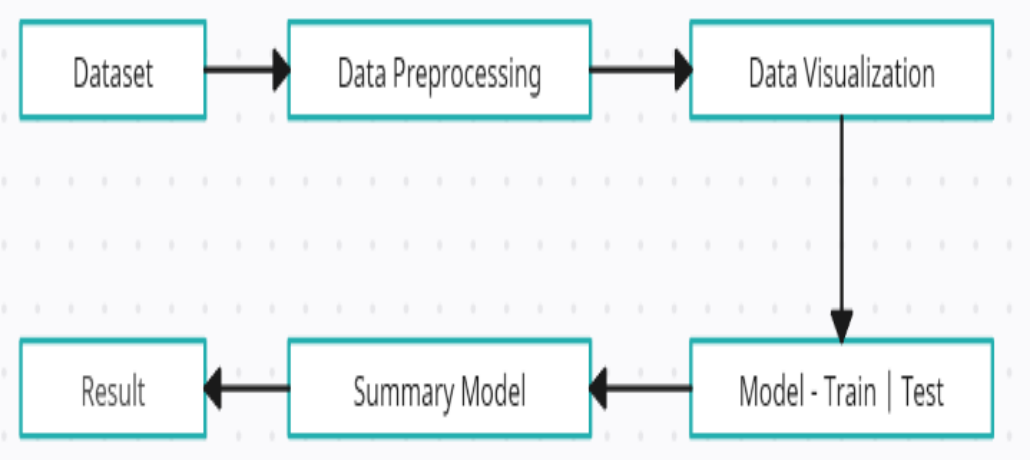

Fig 1: System Architecture

4.RESULTS AND DISCUSSION

Fig 2:Similarity score

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Input

At least three people were reported to have been killed as the cyclone made landfall on a path that is set to cross the world's biggest refugee camp in Bangladesh, home to about a million Rohingyas who fled there years ago from neighbouring Myanmar. Bangladesh and Myanmar evacuated hundreds of thousands of people on fears of widespread destruction. "Cyclone Nargis in 2008 attained a maximum wind speed of 215 kmph. It was the worst meteorological disaster to hit Myanmar.

Output(*Summarised text)*

Cyclone Mocha now has at "Cyclone Nargis in 2008 attained a maximum wind speed of 215 kmph. At least three people were reported to have been killed as the cyclone made landfall on a path that is set to cross the world's biggest refugee camp in Bangladesh, home to about a million Rohingyas who fled there years ago from neighbouring Myanmar. Bangladesh and Myanmar evacuated hundreds of thousands of people on fears of widespread destruction. It was the worst meteorological disaster to hit Myanmar".

Bleu score

The BLEU score evaluates how well a candidate's sentence matches the list of reference sentences by comparing it to one or more reference sentences. It assigns an output score ranging from 0 to 1.

```
BLEUscore : 0.18458827434139816
```

Fig 3. Blue score

5.CONCLUSION

Our model has condensed the information into a single, readily readable paragraph by highlighting the important details and presenting them briefly. The original text must also be read because it might have omitted some context that a human reader would have noticed. Because of this, even while our methodology can save time and effort when summarising a lot of content, it cannot replace a reader's capacity for critical thought and information analysis. Combining human intellect with machine learning is essential for completely understanding complex information.

REFERENCES

1. J. Leskovec, M. Grobelnik, and N. Milic-Frayling, "Learning
2. substructures of document semantic graphs for document summarization," in Proceedings of the KDD 2004 Workshop on Link Analysis and Group Detection (LinkKDD), Seattle, WA, USA, 2004.
3. N. Munot and P.-N. P. I. Sharvari S. Govilkar Department of Computer Engineering, Mumbai University, "Conceptual framework for abstract text summarization," International Journal on

Natural Language Computing (IJNLC) Vol. 4, No.1, February 2015, 2015.

4. E. Lloret, E. Boldrini, T. Vodolazova, P. MartÃnez-Barco, R. MuÃ´soz, and M. Palomar, "A novel concept-level approach for ultraconcise opinion summarization," Expert Systems with Applications, vol. 42, no. 20, pp. 7148 – 7156, 2015. [Online].

5. A. Khan, N. Salim, and Y. Jaya Kumar, "A framework for multi document abstractive summarization based on semantic role labelling," Appl. Soft Comput., vol. 30, no. C, pp. 737–747, May 2015.[Online]

6. A. Khan, N. Salim, and Y. Kumar, "Genetic semantic graph approach for Multi-document abstractive summarization," in Digital Information Processing and Communications (ICD IPC), 2015 Fifth International Conference on, Oct 2015, pp. 173–181

AUTHOR PROFILES

Ms. M. ANITHA completed her Master of Computer Applications and Masters of Technology. Currently working as an Assistant professor in the Department of Masters of Computer Applications in the SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. Her area of interest includes Machine Learning with Python and DBMS.

Ms. V.Raja Rajeswari is an MCA student in the Department of Computer Applications at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. She has a Completed Degree in B.Sc.(computers) from Svl kranthi Degree college Avanigadda. Her areas of interest are DBMS, Java Script, and Machine Learning with Python.