INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 250-253

NLP FOR TEXT SUMMARIZATION

Mr. Aryan singh[1], Mr. Raghvendra singh[2]

[1,2]BCA Department, SRMCM, Lucknow, UP, India.

ABSTRACT

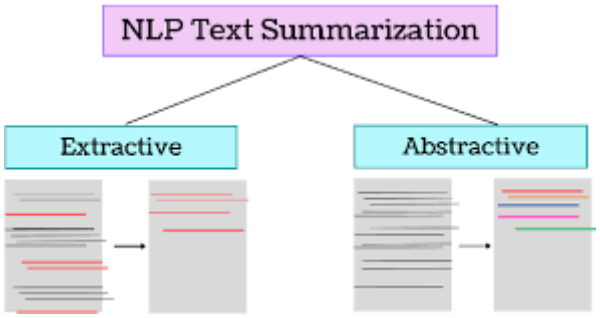
In this study, we investigate different approaches and techniques used in the text summarization sector, from conventional methods to cutting-edge deep learning models. We give a summary of extractive and abstractive summarizing techniques, emphasizing the benefits and drawbacks of each. We also go over evaluation measures that are frequently used to rate the quality of summaries produced by various algorithms. In addition, we introduce new developments in neural network topologies, including Transformer-based models, which have made a substantial impact on abstractive summarization progress. In conclusion, we address possible future paths and obstacles in the field of text summarizing, such as incorporating domain-specific expertise and creating more comprehensible and manageable summarization systems. This thorough summary is an invaluable tool for scholars and professionals who are interested in raising the bar in the field of text summarization and comprehending its terrain.

1. INTRODUCTION

We provide a text summarization system in this research. The suggested approach is predicated on using both morphological features and semantic data to summarize the content from the internet. People have less time to read text data due to its rising length. A system is needed to generate simpler and more concise forms of data because the internet, media, and other data sources provide massive amounts of data. Therefore, a tool that would make reading the complete text or subject easier for people is needed. Users would benefit from and save a ton of time with such systems or solutions. It was not possible for everyone to read and obtain the news material due to hectic schedules.

To be effective, information must be simpler and more dependable.Summaries allow people to quickly make wise selections. The goal is to develop a tool that is effective and generates summaries on its own.

In the field of automated cognition known as "natural language processing," or NLP, computers analyze, comprehend, and derive meaning from human language in a beautiful and practical way. Through the use of implied natural language processing (NLP), designers are able to structure and construct data in order to do activities such as named element acceptance, program rundown, interpretation, relationship generation, judgment investigation, discourse acceptance, and point subdivision.



Workflow: Step in preparation: One procedure that comes before translation is pre-processing. The document has to be moved into a bag of words or phrases. Natural Language Processing (NLP) stages like phrase segmentation, tokenization, stop word removal, and WORKFLOW stemming are included in the pre-processing step. Once pre-processing is finished, word frequency and reverse document frequency values are calculated for each token.

Segmenting sentences: The technique of breaking up a string of written language into its unit or module sentences is known as sentence segmentation.

Tokenization: Tokenization is the process of Step in preparation: One procedure that comes before translation is pre-processing. An entity or collection of interconnected documents is sent into the summarizer system. The document has to be moved into a bag of words or phrases. Natural Language Processing (NLP) stages like phrase segmentation, tokenization, stop word removal, and WORKFLOW stemming are included in the pre-processing step. Once pre-processing is finished, word frequency and reverse document frequency values are calculated for each token.

Segmenting sentences: The technique of breaking up a string of written language into its unit or module sentences is known as sentence segmentation. punctuation is used in languages like English and various others;

Euclidean distances, and the squared errors are optimized by mean. k-means clustering minimizes intra-cluster differences.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 250-253

Stemming: Reducing operationally linked or purposefully connected word forms to their stem, common base, or root forms—typically a written word form—may aid in broadening the scope of Natural Language Processing (NLP) tools. This process is known as stemming.

2. PROPOSED SYSTEM

Automating document summarization is the goal. The suggested system is based on an extractive summarization methodology. The algorithm determines the frequency weight age of every word in every sentence throughout the document, verifies the words' parts of speech, and then assigns a total score to each sentence.

Benefit: The final step generates output based on the highest scoring clusters, which results in insightful and effective summaries. In the clustering phase, the clusters are created based on the sentence scores and are divided into lowest and highest weighted sentences.

K-means The goal of clustering, a quantization vector technique originally used in signal processing, is to separate and observe into k groups where the cluster belongs serving as a prototype for the cluster by matching each observation with the nearest mean cluster centroid or cluster centers. This causes the data space to be divided into Voronoi cells as a result. The regular Euclidean distances would have the most difficult Weber problems; therefore, we squared the Euclidean distances, and the squared errors are optimized by mean. k-means clustering minimizes intra-cluster differences.

For our project, we used the extractive technique for text summarizing. In particular, we have employed the TF-IDF method to summarize. The number of clusters must be known in advance for these clustering-based methods to work, and the MMR approaches contain unknowns regarding coverage and non-redundancy characteristics in the summary, among other things. A comprehensive model that would include an abstract representation for content selection is absent from the Tree Based Method. Method Based on Templates requires creating templates, yet it is too complicated to generalize a template. Method Based on Ontology This method is exclusive to Chinese news sources. Moreover, developing a rule-based system to manage uncertainty is a challenging undertaking.

3. SYSTEM METHODOLOGY

The process of extractive text summarizing involves the following steps:

I. Text Pre-Processing: During this stage, stop words, extraneous letters, and punctuation are removed from the user's input text.

II. Sentence Scoring: This method determines the importance of a sentence.

III. Sentence Selection: The sentences that are prioritized the most will be selected. The sentence will also be shorter in length.

I. Post-Processing: This stage consists of fixing grammatical faults in the text and combining sentences that have the same meaning.

Testing:

- Short Input - When testing with smaller inputs, we receive an error of minimum value indicating that the frequency of the word is not higher than what is needed to compute the summary.
- Foreign Language – It successfully completes the summarization process and produces a meaningful summary when input is provided in any language.
- Incorrect URL – The web scraper cannot obtain the precise data from the URL from which our summary may be made, so it displays the error as shown below if the provided URL lacks defined and sequential data that can be summarized.
- Illogical material: A summary cannot be produced if any illogical or meaningful material is provided as input because it is not logical to do so.

4. RESULTS

Users have access to a convenient platform for summarizing text inputs and viewing both the original and summarized texts thanks to the text summarization website built with Flask. The website is divided into two sections: visitors can enter text in the first textbox on the page and then, after submitting, are taken to the second page, which displays the original and summary texts with their corresponding word counts.

Key features and functionalities:

- Input Interface
- Text Summarization
- Word Count Display

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor: 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 250-253

```
Length of original text: 454
Length of summary text: 196
```

Fig 1.1: Words displayed

5. CONCLUSION

It has been demonstrated that text summaries are helpful for tasks involving natural language processing, such question and answer sessions, as well as for related computer science domains like text categorization and data retrieval. Additionally, search times for information will be more accessible. Sequencing amplifies the effect while its algorithms exhibit less bias than those of human creams. Commercial capture services enable users to handle more texts by providing a text summary system.

News reporters, scholars, and students will find this project useful. Students can use this tool to understand and communicate the content of a research paper or any news article by summarizing it.With the use of this technology, students may read and comprehend any subject more fully.

This tool allows researchers to examine any research project they are working on. There are two ways to summarize text: abstractly and extractively. The process of extractive summarizing is taking the text out of a lengthy piece without altering its meaning. Abstractive summarization, on the other hand, will create an overview of the article.

6. Future Scope:

NLP text summarizing offers a wide range of applications across numerous industries. It can be used by news writers to create the most recent headlines that include succinct summaries. It will mostly be utilized by researchers in the research sector to examine and comprehend intricate scientific publications. This text summarization can also be used by chatbots to give people accurate information. Marketing will make advantage of text summary. Additionally, if someone wishes to purchase a product, the product review and description need to be accurate enough to provide accurate product information.

To summarise, the text summarizer will produce a summary while preserving the meaning of the extensive text. Students, researchers, news reporters, and other users will find it more beneficial.. Text summary will become commonplace because modern publications tend to contain extraneous information. One important tool for making text shorter, easier to read, and more comprehensible is a text summarizer.

7. REFERENCES

[1] T. Al-Taani, "Automatic text summarization approaches," 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), 2017,

[2] P. Patil, S. Dalmia, S. Abu Ayub Ansari, T. Aul and V. Bhatnagar, "Automatic text summarizer," 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2014, pp. 1530-1534, doi:10.1109/ICACCI.2014.6968629.

[3] S. Biswas, R. Rautray, R. Dash andR. Dash, "Text Summarization: A Review," 2018 2nd International Conference on Data Science and Business Analytics (ICDSBA), 2018,

[4] N. Andhale and L. A. Bewoor, "An overview of Text Summarization techniques," 2016 International Conferenceon Computing Communication Control and automation (ICCUBEA), 2016, pp. 1-

[5] P. Janjanam and C. P. Reddy, "Text Summarization: An Essential Study," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-6doi:10.1109/ICCIDS.2019.8862030.

[6] I. Awasthi, K. Gupta, P. S. Bhogal, S. S. Anand and P. K. Soni, "Natural Language Processing (NLP) based Text Summarization - A Survey," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 1310-1317, doi:10.1109/ICICT50816.2021.9358703.

[7] H. T. Le and T. M. Le, "An approach to abstractive text summarization," 2013 International Conference on Soft Computing and Pattern Recognition (SoCPaR), 2013, pp. 371-376, doi:10.1109/SOCPAR.2013.7054161.

[8] P. R. Dedhia, H. P. Pachgade, A. P. Malani, N. Raul and M. Naik, "Study on Abstractive Text Summarization Techniques," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-8, doi:10.1109/ic-ETITE47903.2020.087.

[9] R. Boorugu and G. Ramesh, "A Survey on NLP based Text Summarization for Summarizing Product Reviews," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020, pp. 352-356, doi: 10.1109/ICIRCA48905.2020.9183355.

[10] Jain, A. (2019, April 1). Automatic Extractive Text Summarization using TF-IDF. Retrieved from Medium.com: https://medium.com/voice- tech-podcast/automatic-extractive-text- summarization-using-tfidf-3fc9a7b26f5

[11] KS, J. (2007). Automatic summarising: the state of the art. Inf Process Manag 43, 1449-1487.

[12] Kumar, T. (2014). Automatic Text Summarization. Rourkela.

[13] Mayo, M. (2019, November). Getting Started with Automated Text Summarization. Retrieved from https://www.kdnuggets.com/: https://www.kdnuggets.com/2019/11/getting- started-automated-text-summarization.html

[14] Mr. Vikrant Gupta, M. P. (2012). An Statiscal Tool for Multi-Document Summarization. International Journal of Scientific and Research (ISSN 2250-3153).

[15] Neelima Bhatia, A. J. (2015). Literature Review on Automatic Text Summarization: Single and Multiple Summarizations. International Journal of Computer Applications, 1-5.

[16] Okumura, H. T. (2009). Text Summarization Model based on the budgeted median problem. Proc. 18th ACM Conf. Inf. Knowledge, 1-4.