Text Summarizing Technique Based On Natural Language Processing for Online Review

1st Sharnit Saha
dept. of Computer Science and Engineering(CSE)
School of Data and Sciences(SDS)
Brac University
Dhaka,Bangladesh
sharnit.saha@g.bracu.ac.bd

2nd Arban Rahman Adib
dept. of Computer Science and Engineering(CSE)
School of Data and Sciences(SDS)
Brac University
Dhaka,Bangladesh
arban.rahman.adib@g.bracu.ac.bd

3rd Md Humaion Kabir Mehedi
dept. of Computer Science and Engineering(CSE)
School of Data and Sciences(SDS)
Brac University
Dhaka,Bangladesh
humaion.kabir.mehedi@g.bracu.ac.bd

4th Md Sabbir Hossain
dept. of Computer Science and Engineering(CSE)
School of Data and Sciences(SDS)
Brac University
Dhaka,Bangladesh
md.sabbir.hossain1@g.bracu.ac.bd

5th MD. Mustakin Alam
dept. of Computer Science and Engineering(CSE)
School of Data and Sciences(SDS)
Brac University
Dhaka,Bangladesh
md.mustakin.alam@g.bracu.ac.bd

6th Annajiat Alim Rasel
dept. of Computer Science and Engineering(CSE)
School of Data and Sciences(SDS)
Brac University
Dhaka,Bangladesh
annajiat@gmail.com

***Abstract*—Now-a-days it is quite impossible to imagine life without the internet. Almost everyone now has access to a smart device and the internet because of technological advancements. As the usage of the internet increased rapidly in the past few years, people became dependent on it for various purposes. People have become dependent on the internet for a variety of reasons as its use of it has risen quickly in recent years. From shopping online to booking movie tickets everything is now being done on the internet. But how does one choose the finest product for themselves out of the seemingly endless possibilities available? Well, the solution is quite easy. People frequently read customer reviews of a product before making a purchase. However, occasionally users will post lengthy reviews that are time-consuming and difficult for some people to read. This is where the concept of text summarizing comes into play. This paper discusses a variety of text summarizing methods, ranging from simple to advanced.**

I. Introduction

Extraction of information from many text data as known as text summarizing. Text summary aims to produce a clear, well-organized, and accurate representation of the original texts' content. Numerous industries, including finance, education, marketing, and more, can employ this method. NLP models that have been trained on a sizable corpus of texts may summarize the text. Each year, the usage of the internet has been rapidly rising. The rise in Internet users has resulted in a huge amount of information being saved online every second. There is a need to summarize these facts without sacrificing their original significance. Text summarizing comes into play to ensure that the uniqueness of information is preserved in the summarized version of an actual text. People currently spend most of their time on the internet. There are so many alternatives present on the Internet that it is easy to become confused. People frequently rely on the reviews of others to choose which product or service is best suited to them. It might be an individual's opinion on how a certain restaurant's cuisine tastes or a description of an area they recently visited. The description or review may consist of just a single word or a whole paragraph. Sometimes the description is excessively lengthy, requiring a thorough reading to get the genuine cause. This is where text summary plays its role since it summarizes the whole review while maintaining its originality.

There are multiple types of text summarizing. Based on input there are two types:

Single document Text Summarizing: The input length in this kind of summarizing is brief. The input for Summarizing will only consist of one document. This was employed in the early stages of text summarizing.

Multi-document Text Summarizing: This approach involves providing many papers as input for a summarizing technique since the length of the input on a particular topic is too long. The challenge in this situation is that the contents of many texts may vary. An effective summary strategy frequently condenses the major ideas while preserving readability, and

completeness, and without omitting the crucial sentences.
Also, based on output there are two types:
Extractive Text Summarizing: The technique of extractive text summarizing involves selecting phrases from the entire text that could convey the same meaning as the entire text but in a more reduced form. Most text summary methods in use today are of the extractive variety.
Abstractive Text Summarizing: This more sophisticated method of text summary entails creating phrases or sentences that don't belong in the text but nonetheless convey the meaning of the entire text. Although the model finds it more challenging to construct phrases or sentences that could convey the same information, this way is more captivating.

II. Literature Review

There are still numerous notable works in text summarizing from the last several years. Most earlier studies mostly focused on text summaries of single documents, however as technology has evolved, computational power has also significantly increased, leading to faster processing of documents and greater accuracy than prior methods.
In the year 2012, an extractive-based Text Summarizing technique using a Genetic algorithm was proposed by Niladri Chatterjee, Shubham Goyal, and Amol Mittal [1]. They displayed the single document as a Directed-Acyclic-Graph in their paper. Each edge of the DAG was given a weight based on a schema that is defined in the paper. They apply an Objective function to represent the summary's standard in words like readability, the degree to which sentences are connected to one another, and the topic relation factor. The goal of the Genetic Algorithm is to maximize the Objective function by picking the most important sentences from the entire text. The first step is to determine how closely connected the sentences are to one another. The Topic Relation Factor, which is calculated next, recommends giving the most weight to sentences that are most related to the input query. The Objective Function of the summary can be calculated when the criteria have been considered.
Word vector embedding was the method utilized by Aditya Jain [2]. in their model for extractive text summarizing. According to their paper, there are four main issues to address while obtaining information. The most important sentences from the text are being highlighted, extraneous information that is irrelevant to the theme of the document is being removed, details are being minimized, and the significant information that was initially retrieved is being combined into a streamlined and ordered report. They suggested a Word Vector Embedding methodology to extract the important information, and then they employed a Neural Network for Extractive Summarizing by employing the Supervised Learning method. The results were more accurate when compared to the earlier summarizing techniques, according to their testing on the DUC2002 dataset. The results were good, but they may be better if the dataset's size and theme variety were expanded, as well as if more effective summarizing techniques like Sequence-to-Sequence Recurrent Neural Networks were used. When a word has many meanings, summarizing, in Shi Ziyan's opinion [3], cannot produce reliable results. Therefore, it is also necessary to have a specific domain understanding of the document's central idea. This highlights the importance of domain-specific text summarizing. But when the referring is done incorrectly, a problem occurs. This work, therefore, provides a co-reference resolution approach to address this issue and produce accurate results.
Priya Pawar [4] talked about how important it is to categorize and summarize product reviews. SVM and Naive Bayes hybrid classifiers were employed. They also concluded that accuracy may be raised along with the number of classifiers.
For the purpose of summarizing online reviews, the Sequence-to-Sequence model may be acceptable as a more accurate and extremely close description of the product review that the other user has submitted for a specific product.

III. Research Objective

Text summarization is a technique for producing a brief, accurate summary of lengthy texts while concentrating on the passages that provide useful information and keeping the overall meaning intact. The goal of automatic text summarization is to reduce lengthy documents to their essential content. Manual text summarization could be time-consuming and expensive. The current outburst of non-structured textual data in the digital sphere necessitates the creation of automatic text summarization tools that make it simple for users to draw conclusions from them. We now have immediate access to vast amounts of information. However, the majority of this data is unnecessary, trivial, and might not convey the intended meaning. For instance, if you are searching for specific information in a news article online, you may need to sift through the article's content and spend a lot of time eliminating the extraneous information before finding what you are looking for. Accordingly, it is increasingly important to use automatic text summarizers that can extract useful information while excluding inessential and irrelevant data. Implementing summarization can make documents easier to read, cut down on the time spent looking for information, and make it possible to fit more information into a given space.

IV. Methodology

The process of condensing long texts into manageable paragraphs or sentences is called NLP text summarization. This technique preserves the text's meaning while also extracting important information. This shortens the amount of time needed to comprehend lengthy items like articles without omitting important details.
Extractive text summarization, as the name implies, gathers important information from the copious amounts of text given and summarizes it in an understandable manner. The method is very simple because it extracts texts based on criteria like the text that needs to be summarized, the Top K sentences, and the importance of each sentence to the overall subject. However, this also means that the method is constrained by predetermined parameters, which could, in some circumstances,

bias the extracted text. The most popular technique used by automatic text summarizers is extractive text summarization because it is straightforward in most use cases.

Abstractive text summarization produces understandable sentences from the entire text that is given. Large amounts of text are rewritten by it by producing palatable representations, which are then processed and summarized by natural language processing. The ability of this approach to process text and use NLP to smooth out the kinks is almost AI-like, which makes it stand out from other approaches. Abstract summarization is often much more useful than the extractive method, even though it may not be as straightforward to use. It resembles fully developed AI writing tools in many ways. This does not negate the need for extractive summarization, though.

Here in this technique, we will combine a large number of articles in a short summary. To get this result we have to follow some steps to get a good summary. The NLP helps us to get this method. Process the article's language then combine it with the near words to summarize the whole article. So we can be able to get the most possible accurate result. Here is the diagram of the process where we can see how to extract the text, sentence and then combined it in a summary. (Fig 1)

Fig. 1. Process of Summarizing

In this approach, we have created a technique where we can see that the article is first combined then converted into text then split into different sentences. After that, the sentence will start to reform in a vector form. The vector form then finds a similar matrix according to its own algorithm. After finding a similar matrix it will approach the sentence ranking the sentence will be ranked according to the matched combination of the article. Lastly, it will provide a summary of the article. Since important sentences in the text are those that would be connected to more other sentences than not, we discover that process end up representing the topics covered in the text (if you consider sentences as vertices and sentence similarity represented by edges) The same techniques can be used for many different languages, so language-specific processing is not necessary. We frequently discover that the semantic knowledge gained through sentence similarity improves summarization performance beyond more straightforward frequency approaches.

Conclusion

The summary cuts down on reading time. Summaries simplify the selection process when researching documents. The efficiency of indexing is increased by automatic summarization. Algorithms for automatic summarization are less biased than human summarization. In question-answering systems, personalized summaries are helpful because they offer individualized information. Commercial abstract services are able to process a greater volume of text documents by utilizing automatic or semi-automatic summarization systems. Text summaries are Immediately efficient, It requires time and effort to read the entire article, analyze it, and extract the key ideas from the written text. A 500-word article takes at least 15 minutes to read. Automatic summary software summarises texts of 500–5000 words in a fraction of a second. This enables the user to read less information while still obtaining the most important details and making reliable decisions. Since they are built on linguistic models, summarizers can automatically summarize texts in a variety of languages, including English and Russian. They are therefore ideal for those who read and use multilingual information. This process can be helped with summaries of any article for a fast process. Thus research helps to fill up those gaps in the online review part. So anyone can get any kind of summary of the online review.

References

[1] Niladri Chatterjee, Amol Mittal and Shubham Goyal's "Single Document Extractive Text Summarization Using Genetic Algorithms" (2012)

[2] Aditya Jain, Divij Bhatia, Manish K Thakur's "Extractive Text Summarization using Word Vector Embedding" (2017).

[3] Shi Ziyan "The Design and Implementation of Domain-specific Text Summarization System based on Co-reference Resolution Algorithm" 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery.

[4] Priya Pawar, Siddhesha Tandel, Shweta Bore Nikita Patil "Online Product Review Summarization" 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIEC).

[5] Shivangi Modi Rachana Oza "Review on Abstractive Text Summarization Techniques (ATST) for single and multi documents" 2018 International Conference on Computing, Power and Communication Technologies.