IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

EXTRACTIVE SUMMARIZATION

Rajul Morande, Sanjivani Revshette, Sunanda Gandhi, Onkar Sonavane, Amruta Aphale
Student, Student, Student, Student, Assistant Professor
Department of Artificial Intelligence and Data Science
AISSMS Institute of Information Technology, Pune – 411001.

Abstract: A concise and clear summary of the original text is produced by text summarising. Without leaving out any essential information, the summary emphasises the most intriguing parts of the original text. Applications like Inshort and Blinklist, which not only save time, but also effort, are widely available and feature news summaries. The process of manually summarising a text can take some time. Thankfully, the technique can be mechanised with the use of algorithms.

INTRODUCTION

Use non-convex entropy minimization to summarise Amazon product reviews. The main concept is to choose a small sample of reviews that accurately describe the attributes of the product. This can be accomplished by using a set of properly thought-out criteria to search for a sparse representation of the complete dataset in a suitable dictionary. Consequently, the chosen representatives conform to this sparse representation. It is effective to locate such representation via nonconvex entropy minimization.

Extractive and abstract accounts in speech are two basic arrangements of information summarization. Extractive writing is the process of selecting the ultimate useful dispute or phrase from the original theme and bearing a brief dossier. This is in comparison to abstractive summarization, which is cultivating new sentences that precisely show the elements of the original document.

DEMAND FOR NATURAL LANGUAGE PROCESSING

Scaling other tasks involving language is made possible by natural language processing, which enables computers to speak with people in their own language. NLP, for instance, enables computers to read text, hear voice, analyse it, gauge sentiment, and ascertain which bits are crucial.

The complexity and variety of human language is remarkable. We have several methods to communicate, both orally and in writing. There are many different languages and dialects, and each one has its own set of grammatical and syntactical conventions as well as words and slang. We frequently delete punctuation, misspell words, or shorten sentences when we write. We talk with regional accents, mumble, stammer, and use words from different languages.

While supervised, unsupervised, and deep learning are increasingly often used to mimic human language, syntactic and semantic knowledge as well as domain knowledge are still necessary and not always available in these machine learning techniques. NLP is significant because it offers valuable quantitative structure to the data for many downstream applications, such as speech recognition or text analytics, and assists in resolving linguistic ambiguity.

For text summary, there are essentially two different methods:

1. Extractive Summarization
2. Abstracting and summing up

EXTRACTIVE SUMMARIZATION

We select only the crucial phrases or sentences from the original text and extract them. Extractive summarization refers to the process of creating summaries that are solely composed of content that has been taken from the source text. The typical issue that emerged from the extractive summarization study at first was figuring out where the sentence was in the text and how frequently certain phrases appeared. The Information Extraction (IE) technique was used in the following experiment to address the extraction issue and create a summary with more precise results and greater accuracy.

Contrary to extractive summarising, sentences produced by abstractive summaries are new sentences, or what is known as paraphrases, which construct summaries using terms that are not in the text. Because they involve considerable natural language processing, abstractive summaries are much more complicated and challenging to create than extractive summaries. The linguistic approach and the semantic approach are the two categories into which approach strategies in abstractive summarization are often divided.

APPROACH FOR EXTRACTIVE TEXT SUMMARIZATION(ETS)

The most significant sentences and paragraphs from the provided material are chosen in extractive text summarization to create the summary. The summarising procedure in ETS entails the following four steps:

(a) Stop words Removal : There are a lot of stop-words in the provided input document that shouldn't be used in the text summarising.

(b) Segmentation : Segmentation separates the input document into its constituent paragraphs, phrases, and words.

(c) Tokenization : Segmentation and tokenization are related concepts. It separates the words into unigrams, bigrams, trigrams, etc. Bigram words combine two words whereas unigrams only have a single word tokenization. Similar definitions apply to trigrams and four-gram words.

(d) Stemmatization of words : A word can have many tenses, plural and single variants, and POS tags. Only root form is considered.

The creation of the summary, which comes last, creates a summary and positions pertinent information from the main content in each sentence.

TECHNIQUES IN EXTRACTIVE SUMMARIZATION

The current era began in the late 1950s. There are several methodologies that have been utilised in ETS research in the past and present, and ETS can be many taxonomies based on NLP summarization. A one-by-one method is provided that may be used to summarise ETS. Unsupervised, semi-supervised, and supervised learning are the three categories under which learning types are categorised. These methods are beneficial for a pertinent summary.

- **Fuzzy logic approach :** This method has been employed by several researchers before. It is based on the proper noun, the primary idea, and a number of anaphors with binary values between 0 and 1 ([0-1]), albeit it isn't always precise. To deal with uncertainty in an unsupervised way, the fuzzy logic model has to incorporate common sense thinking.

- **Graphical based approach :** In this method, the graph is constructed using a collection of produced edges and vertices from the document. Sentences in the document served as the graph's vertices, while words in the document served as its edges.

- **Statistical based approaches :** This method is frequently used in text summarising to provide an immediate, pertinent summary. Because centrality and frequency are its primary applications, the statistical technique is the focus of many academics. The concepts of centrality and frequency are also applied in unsupervised methods. This method solely considered non-linguistic aspects of the document, such as word placement, sentence structure, and many other aspects to the tf-idf. Later, a keyword list is created.

- **Machine learning approach (MLA) :** In labelled data or training datasets, these techniques are also applied. Without training data, its summary is generated using an unsupervised learning approach. Recurrent neural networks (RNN), convolutional neural networks (CNN), RBM, and autoencoder, among other methods, are some of the approaches used in deep learning. Techniques for supervised learning include the Naive Bayes classifier, Genetic Algorithm, SVM, Regression, and Multilayer Neural Network. Using semi-supervised learning approaches, a useful classifier is created by combining unlabeled and labelled data. It employs a variety of strategies, including SVM and the naive Bayes classifier.

ARGUMENTATION

There are only a handful ETS techniques, and they all have their limits. The fuzzy rule is necessary for human specialists, and the fuzzy logic method improves sentence ranking concerns. With the aid of the ranking algorithm, graph methods may be summarised, and the correctness of the summary depends on the use of mathematical functions in the graph approach. Languages were used in the discourse technique to construct a summary, but it lacked consistency and cohesiveness and required a domain-based dataset. The statistical technique does not require a training dataset, but it processes data in the simplest way possible and generates a summary devoid of speech or semantic understanding. Machine learning methods based on features and high number feature tests for the performance of the datasets and its application for the much bigger dataset which can produce the output with greater accuracy.

CONCLUSION

Summarization has a fascinating topic of the research area in natural language processing and thorough study of a number of methods that provide information summaries that are non-redundant, succinct, logical, free of sentence ordering issues, and pertinent.

We will review methods for abstractive text summarization in the future and discuss their taxonomy, benefits, and drawbacks. We will learn more about the problems and research concerns associated with abstractive text summarization.

To do this, we expand a summarization method that was previously given by substituting a wide range of linguistic annotation types for bigrams, including as n-grams, verb stems, frames, ideas, chunks, connotation frames, entity types, and discourse connection sense-types. To assess the capacities of information significance detection, we provide two innovative assessment methodologies. When source document sentences must be rated in our trials, bigrams perform the best overall. These findings validate the usage of bigrams in summarization systems by the creators of those systems.

ACKNOWLEDGMENT

We appreciate Ms. Amruta Aphale, our guide, for her unwavering support and direction. We were able to effectively complete our job thanks to her coaching. Additionally, we would like to express our gratitude to the institution, division, and specific people for their unwavering support and direction during this study. The success of this initiative would not have been achieved without their assistance.

REFERENCES

[1] Tao, Y., Zhou, S., Lam, W., Guan, J. (2008). Towards more effective text summarization based on textual association networks. In Semantics, Knowledge and Grid, 2008. SKG'08. Fourth International Conference on, pp. 235-240. https://doi.org/10.1109/SKG.2008.17

[2] "Extractive Text Summarization Using Recent Approaches: A Survey" by Avaneesh Kumar Yadav* | Ashish Kumar Maurya Ranvijay | Rama Shankar Yadav [3] https://en.wikipedia.org/wiki/Feature_extraction

[3] "Comprehensive Guide to Text Summarization using Deep Learning in Python" Download Share Aravindpai Pai — Published On June 10, 2019 and Last Modified On May 10th, 2020

[4] L. Abualigah, M.Q. Bashabsheh, H. Alabool, M. Shehab "Text summarization: A brief review"