ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 7 | Issue: 4 | April 2021|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2021: 8.047 || ISI Value: 1.188

AUTOMATIC TEXT SUMMARIZATION USING DEEP LEARNING

Keerthana P[1]
[1]PG Student,
Department of Computer Science,
Dr.N.G.P Arts and Science College,
Tamil Nadu, India

ABSTRACT

Text summarization is a process of extracting collecting important information from original text and present the information in the form of summary. Text summarization has become the necessity of many applications for example search engine, business analysis, market review. Summarization helps to gain required information in less time. This paper is an attempt to summarize and present the view of text summarization from every aspect from its beginning till date. The two major approaches i.e., extractive and abstractive summarization is discussed in detail. The technique deployed for summarization ranges from structured to linguistic. In Indian many languages also the work has being done, but presently they are in infancy state. This paper provides an abstract view of the present scenario of research work for text summarization. In an computerized machine of summarization which is wreck up into the subsequent steps: Pre-processing (sentence segmentation, tokenization, give up words removal), Feature Extraction, Sentence Scoring, Sentence Ranking and Summary Extraction. Finally, this paper collects the most crucial today's and applicable lookup interior the zone of the textual content summarization to assessment and assessment for future research.

KEYWORDS - ***Text Summarization, extractive and abstractive summarization.***

I.INTRODUCTION

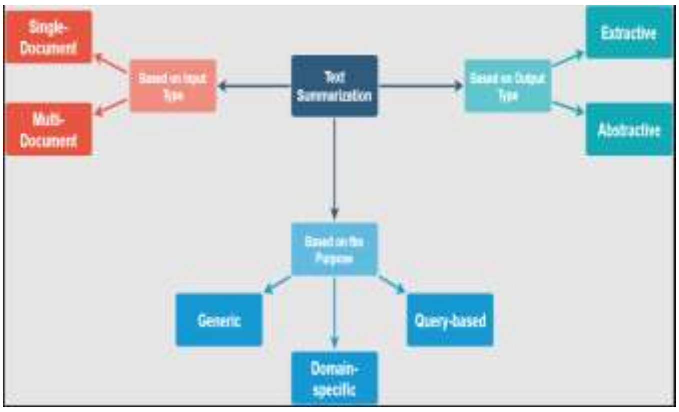
With increasing amount of knowledge it becomes extra and greater challenging customers to search efficaciously for unique content to understand a influential, necessary and applicable material today's information technology wide variety of humans is checking out informative on web, each time it's impossible that they thought to get all application records in single document ,or on a single website. They might get variety of sites as a search result [5]. This has given the new solution that is associated to records processing and computing device mastering which returns query unique Information from massive set of offline archives and represents as one file to the user. So, computerized summarization is a fundamental location in Natural Language Processing (NLP) research. Automated summarization affords single file summarization and multi-document summarization [3].

II. TYPES OF SUMMARIZATION

A. Multi-Document Merger

The merging of data from multiple documents is called multi-document merger. Data is found in unstructured or structured form and many times we have to generate summary from multiple files in less time, so, multi-document merger technique is useful. Multi-document summarization generates information reports that are both concise and comprehensive. With different opinions being put together, every topic is described from multiple perspectives within a single document. The goal of a brief summary is to simplify information search and save the time by pointing to the most relevant information.

Fig1: Unsupervised text summarization



B. Extractive Text Summarization

Extractive summarizer aims at selecting the foremost Relevant sentences within the document whereas maintaining a reduced redundancy within the outline. It is created by reusing portion (word, sentences etc.) of input text verbatim .Example: Search engines typically generate extractive summaries from web pages.

1. Term Frequency Inverse Document Frequency (TFIDF) approach:

Bag words model is made at sentence level, with the traditional term frequency and sentence frequency algorithms, wherever sentence frequency is that the range of sentences within the document that have that term, words that occur frequently within the documents is additionally taken because the question words.

2. Clustering based approach

Documents area unit consist of mistreatment term frequency and inverse document frequency (TFIDF) of various Extractive summarizer aims at selecting the foremost relevant sentences within the document whereas maintaining a reduced redundancy within the outline. It is created by reusing portion (word, sentences etc.) of input text verbatim. Example: Search engines typically generate extractive summaries from web pages words. Term frequency in this context is that the average range kind of document over of existences of similar the cluster

C. Abstractive Text Summarization

The rule based method comprises of three step Firstly,The documents to be classified are represented in terms of the rule based method Their categories. The categories can be from various Domains. Hence the first task is to sort these. The next thing is to make questions supported these categories amongst the varied categories like attacks, disasters, health etc. The context selection module selects the best candidate amongst these.-Generation patterns are the Methods employ more powerful natural language processing

1. Rule Based Method

Techniques to interpret text and generate new summary text,As opposed to selecting the most representative existing excepts to perform the summarization. A) In this method,information form source text re-pharased. but it is harder to use because it provides allied problems such as semantic Representations.

Example: Book Reviews-if we want a summary of book the lord of The Rings then by using this method we can make summary from it. The used for the generation of summary sentences.

III.LITERATURE SURVEY

An improved method of automatic text summarization for web contents using lexical chain with semantic-related terms proposes an improved extractive text summarization method for documents by enhancing the conventional lexical chain method to produce better relevant information. Then, Author firstly investigated the approaches to extract sentences from the document(s) based on the distribution of lexical chains then built a transition probability distribution generator (TPDG) for n-gram keywords which learns the characteristics of the assigned keywords from the training data set. A new method of automatic keyword extraction also featured in the system based on the Markov chains process. Among the extracted n-gram keywords, only unigrams are selected to construct the lexical chain [1].

In paper author first extracted multiple candidate summaries by proposing several schemes for improving the upper-bound quality of the summaries. Extensive experiments have been conducted on a benchmark dataset [2].

Automatic text summarization within big data framework demonstrates how to process large data sets in parallel to address the volume problems associated with big data and generate summary using sentence ranking. TF-IDF is used for document feature extraction. Map Reduce and Hadoop is used to process big data [3].

Extractive documents summarization based on hierarchical GRU proposes two stage structure 1) Key sentence extraction using Levenshtein distance formula 2) Recurrent neural network for summarization of documents. In extraction phase system conceives a hybrid sentence similarity measure by combining sentence vector and Levenshtein distance and integrates into graph model to extract key sentences. In the second phase it constructs GRU as basic block, and put the representation of entire document based on LDA as a feature to support summarization [4].

Extractive algorithm of English text summarization for English teaching is based on semantic association rules. In this paper relative features are mined among English text phrases and sentences, the semantic relevance analysis and feature extractions of keywords in English abstracts are realized [5].

IV.PROPOSED METHODOLOGY

In above literature survey we found that all summarization frameworks are unique in their own way with respect to document processing, algorithms and final outputs. To overcome the limitations discussed above for existing systems, we suggest following methods.

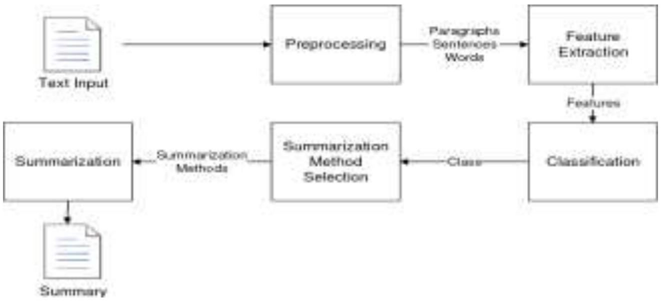


Fig2 Extractive Summarization

1. Clustering with cosine similarity algorithm for sentence extraction

Previously we analyzed some limitations of existing systems one of them was single domain summarization that is Algorithm only works on specific documents like scientific journals', sports, news documents. To avoid this we suppose to use cosine similarity algorithm which gives better sentence extraction result regardless of the type of document or size of the document. While extracting sentences we will treat a heading as a general sentence so the system will perform on document with or without heading.

1. The NEWSUM algorithm for generating clusters

To increase accuracy we have to use clustering so that we can avoid unrelated documents, on top of that both algorithms have minimum time complexity which will help to minimize overall system execution time.

2 .Position score algorithm to rank the sentences

To rank the extracted sentences we use position score Algorithm. It helps to maximize the accuracy rate of the system.

V.CONCLUSION

The day the growth of data is increased in structured or unstructured form and we need summary from that data in less time. So there is a need for automatic text summarization tool. In this survey paper we have discussed about various types of text summarization techniques. Further, limitations throughout found the papers are discussed and probable solutions are also given. To overcome the drawback of existing models, here we have proposed a new model. It includes clustering with cosine similarity algorithm, the NEWSUM algorithm and position score algorithm. The proposed framework is under development. The presently taken results are giving positive outcome from proposed system.

VI.REFERENCES

1. *HtetMyet Lynn 1 , Chang Choi 2 , Panko Kim"An improved method of automatic text summarization for web contents using lexical chain with semanticrelated terms", SpringerVerlag Berli Heidelberg 2017*
2. *Xiaojun Wan 1 , FuliLuo 2 , Xue Sun Songfang Huang3 , Jinge Yao "Crosslanguage document summarization via extraction and ranking of multiple summaries" Springer Verlag London 2018.*
3. *Andrew Mickey and Israel Cuevas "AUTOMATIC TEXT SUMMARIZATION WITHIN BIG DATA FRAMEWORKS", ACM 2018.*

4. *Yong Zhang, Jinzhi Liao, Jiyuyang Tang "Extractive Document Summarization based on hierarchical GRU", International Conference on Robots & Intelligent System IEEE 2018.*
5. *Lili Wan "Extractive Algorithm of English Text Summarization for English Teaching" IEEE 2018.*
6. *Anurag Shandilya, Kripabandhu Ghosh, Saptarshi Ghosh "Fairness of Extractive Text Summarization", ACM 2018.*
7. *.P.Krishnaveni, Dr. S. R. Balasundaram "Automatic Text Summarization by Local Scoring and Ranking for Improving Coherence", Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication.*
8. *Bagalkotkar, A., Kandelwal, A., Pandey, S., &Kamath, S. (2013, August). "A Novel Technique for Efficient Text Document Summarization as a Service", In Advances in Computing and Communications (ICACC), 2013 Third International Conference on (pp. 5053). IEEE.*