Artificial Intelligence Review (2021) 54:5897–5929
https://doi.org/10.1007/s10462-021-09964-4

Study of automatic text summarization approaches in different languages

Yogesh Kumar[1] · **Komalpreet Kaur**[2] · **Sukhpreet Kaur**[3]

Accepted: 21 January 2021 / Published online: 12 February 2021

Abstract
Nowadays we see huge amount of information is available on both, online and offline sources. For single topic we see hundreds of articles are available, containing vast amount of information about it. It is really a difficult task to manually extract the useful information from them. To solve this problem, automatic text summarization systems are developed. Text summarization is a process of extracting useful information from large documents and compressing them into short summary preserving all important content. This survey paper hand out a broad overview on the work done in the field of automatic text summarization in different languages using various text summarization approaches. The focal centre of this survey paper is to present the research done on text summarization on Indian languages such as, Hindi, Punjabi, Bengali, Malayalam, Kannada, Tamil, Marathi, Assamese, Konkani, Nepali, Odia, Sanskrit, Sindhi, Telugu and Gujarati and foreign languages such as Arabic, Chinese, Greek, Persian, Turkish, Spanish, Czeh, Rome, Urdu, Indonesia Bhasha and many more. This paper provides the knowledge and useful support to the beginner scientists in this research area by giving a concise view on various feature extraction methods and classification techniques required for different types of text summarization approaches applied on both Indian and non-Indian languages.

Keywords Text summarization · Machine learning · Feature extraction · Classifications · Natural language processing

✉ Yogesh Kumar
yogesh.arora10744@gmail.com

Komalpreet Kaur
chadhakomal18@gmail.com

Sukhpreet Kaur
er.sukhpreetkaur@gmail.com

[1] Department of Computer Science & Engineering, Chandigarh Group of Colleges, Landran, Mohali, India

[2] Department of Electronic & Communications, Punjabi University, Patiala, India

[3] Department of Computer Science & Engineering, Chandigarh Engineering College, Landran, Mohali, India

1 Introduction

A technique to reduce the long sentences into short is known as text summarization. It creates a summary of whole text only including the main points. In other words, we can say it is the distillation of useful text from main text. Main advantages of using this technique are that it reduces reading time, accelerates the procedure of research, and increases the overall area so that more information can fit in. Automatic text summarization is a popular difficulty in machine learning and natural language processing (NLP) (Widyassari et al. 2019). Machine learning models are trained to understand the full data and filter out only the useful information. It is easy to summarize voluminous data through machines; otherwise it becomes difficult, time consuming and expensive if done manually. But then the very first question arises in a mind is about the selection of core content from main document and next is about the compression of selected data. Research in the field of text summarization focuses more on the product, that is, the summary and less on the reasoning base for understanding text (Ranabhat et al. 2019). Certain constraints in existing systems would have an advantage of improved understanding of the cognitive basis of the task. Early investigation in the field of summarization was based on single document summarization, on the other hand, in today's scenario various approaches concentrate on multiple document summarization.

Today most summarization algorithms target at the creation of abstracts given the problems related to automatic generation of attractive content in random fields. In general, there are several factors that define the selection of content from the source document for summary generation, such as, the readers either expert or non-expert, undoubtedly effect the content selection.

As shown in the Fig. 1, the input text document of any language which includes newspaper articles, medical related documents, legal documents and any type of report documents. The primary objective of any ATS system of different language is to produce an automated summary from the input text document or multiple documents such that the summary is shorter in length than the actual document(s) which includes only the most important information and does not contain the duplicate data. There are various approaches to do the Automatic Text Summarization (ATS), but mainly we focus on two approaches, namely, Extraction based summarization and the Abstraction based summarization. Extraction based summarization is simple to implement but abstraction-based summarization is better. Extraction based method extracts all the key points from the main document and makes summary out of it, without changing the original text (Moratanch and Chitrakala 2017). On

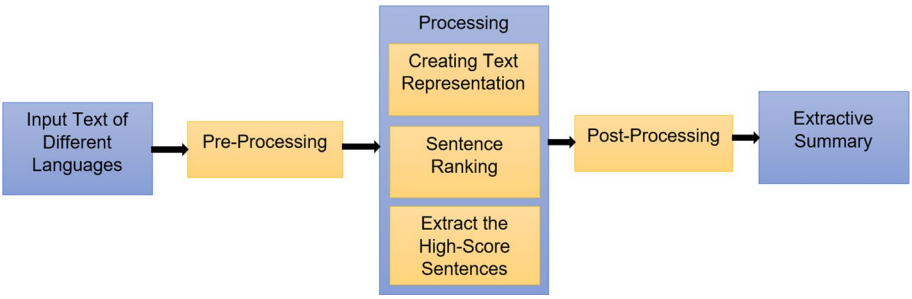


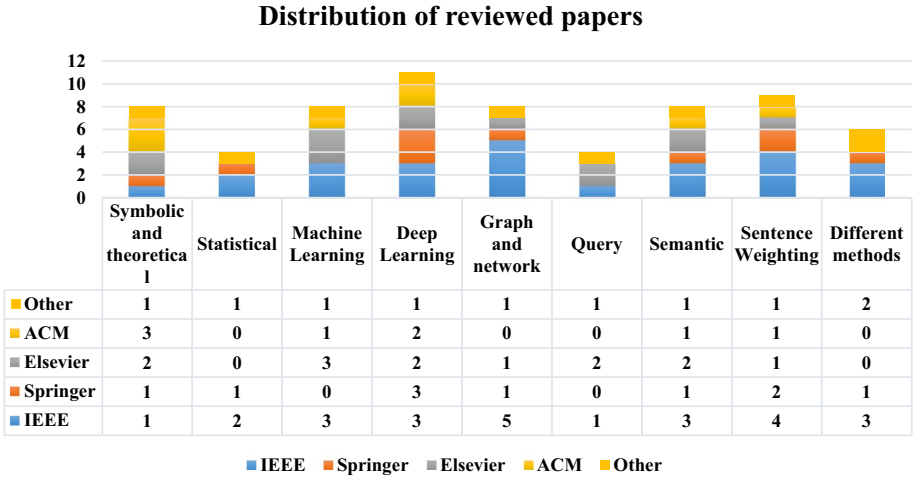
Fig. 1 General structure of Automatic Text Summarization (Widyassari et al. 2019)

the other hand, abstraction-based method does its job just like humans do. It makes changes in the original text and adds phrases to it leading to grammatically correct summary. Hence, abstraction approach carries out better summarization than extraction approach. It is difficult to develop the text summarization algorithms for abstraction approach this is the reason extraction approach is more popular (Bhatia and Jaiswal 2016). Earlier, extractive summarizers generally used sentence scoring for summary extraction. But presently there are numerous new techniques available which utilize linguistic or statistical properties of text for summary generation, for example, standard keyword, high frequency words, cue method, location method, and title method for measuring the weight age of sentences. In this paper, we have reviewed various research articles on selected content and tools used for text summarization along with resources used for evaluation. For apprentice researchers, this work proves to be beneficial in providing the brief idea about various approaches used in text summarization systems along with outcomes achieved for each.

1.1 About the study

The aim of this paper is to provide a comprehensive view on text summarization. This paper takes into account all the details related to text summarization of various Indian and Foreign languages using different approaches. We have seen various research papers in our literature survey related to it. But no single survey is available which contains text summarization of both Indian and foreign languages along with various approaches for the same. This paper is unique, as it contains an elaborate discussion about summarization of various Indian regional and foreign languages and comparative analysis of various approaches and techniques employed for them. This survey article aims at putting forward a systematic and broad review on the automatic text summarization, with the covering main areas, for instance, dataset sources, challenges, benefits and study of various text summarization-based approaches along with feature extraction techniques, classification methods, other approaches and its outcome in terms of various parameters.

The rest of the paper is organised in various sections, firstly brief about text summarization is given in Sect. 1 named introduction that also includes the Fig. 2 for description of all



	Symbolic and theoretical	Statistical	Machine Learning	Deep Learning	Graph and network	Query	Semantic	Sentence Weighting	Different methods
Other	1	1	1	1	1	1	1	1	2
ACM	3	0	1	2	0	0	1	1	0
Elsevier	2	0	3	2	1	2	2	1	0
Springer	1	1	0	3	1	0	1	2	1
IEEE	1	2	3	3	5	1	3	4	3

Fig. 2 Distribution of paper reviewed for text summarization methods

the papers taken from different organised sources. Section 2 gives the novelty in the study and motivation behind the work that mainly highlights the objective of text summarization along with benefits one can get using it. The background study is illustrated in Sect. 3 includes framework for automatic text summarization, benefits, challenges and its application in various domain such as media monitoring, Search market and SEO, Internal document workflow, Internal document workflow, social media marketing and many more. This section also contain brief about various datasets used for the same. Section 4 includes the reported work in which various papers are highlighted in terms of different approaches used for text summarization. Forwarded with Sect. 5 that gives the comparative study of existing work in various Indian and Foreign languages along with the approach, Dataset and parameters associated with their work. In the end Sect. 6 concludes the work that helps researchers in choosing the best area along with approaches applicable in their work for future.

2 Motivation

The objective of text summarization is to present the matter of a document in a compressed form that satisfies the requirements of the reader. Vast amount of data is available on electronic devices and World Wide Web which is genuinely difficult to digest (Nathani et al. 2020). All information sources like News, biographical, historical information etc. hold large amount of data which is difficult to read, so some sort of compression of data is required. In this digital era, the amount of exchange of digital information is really big. So here is a requirement to invent machine learning algorithms that can automatically reduce lengthier scripts and provide precise summaries that can easily pass the actual messages. Following points gives the brief overview about benefits of text summarization (Shimpikar and Govilkar 2017):

- Summaries cut scanning time of any document.
- When exploring research papers, help in sorting and selecting papers.
- ATS enhances the value of indexing.
- ATS algorithms are less partial than man-made generated summaries.
- Custom-made summaries are beneficial in question–answering systems as they deliver personalized data.
- With semi-automatic or automatic summarization methods, help in processing a greater number of texts in commercial abstract services.

Use of ATS is very significant for various users due to saving of time that get wasted in case of manual summarization. Along with this, it also help in getting summary of single or multiple documents related to same topic. The final summary can also give the general overview of various topics in the input documents and in return user can get more details from original documents (Baruah et al. 2020).

3 Background study

In this section, we have discussed various approaches and the procedures used in text summarization, namely, abstractive and extractive approaches. We have also presented structured framework for text summarization, which basically includes three steps; identification, interpretation and generation of summary. We also mentioned about the role machine learning in

text summarization and how it is helping in advancing the performance of summarization tools (Nallapati et al. 2017). Highlighted about the key points to be taken into account while designing any machine learning algorithm for ATS. We also discussed about various benefits of ATS and challenges related to it. ATS has various applications in different domains. We have highlighted about several use cases of automatic text summarization in our work. Data on different languages is required to carry out experiments or test designed models for summary generation (Florescu and Jin 2019). For this purpose, we listed few of the dataset sources for text summarization. In this section we will also discuss about the goals and objectives behind writing this paper.

3.1 Framework for automatic text summarization

The proposed framework have different phases in which sentences are pre-processed, extracted and classified for generating the text summary. The framework focus on the different langauges text document for pre-processing which includes the tokenization, sentence boundary identification, stop word elimination-common words with no semantics and stemming obtained through the radix or root of each word, which highlight its semantics.

In processing step, parameters persuading the relevance of sentences are decided and computed and then weights are allocated to these parameters using weight learning method (Parveen et al. 2016). Using feature-weight equation, final weight of each sentence is determined and highest weighted sentences are used to generate the final summary. Following Fig. 3 shows the steps involved in extractive text summarization.

3.1.1 Pre-processing

The very first step in pre-processing is tokenization, in which the document is broken into smaller units of paragraphs, sentences and words. It is followed up by stop word elimination step in which removal of repeatedly occuring words in text is done (Mihalcea and Tarau +). The next step is stemming, in which derived words are converted back into their stem or root words.

3.1.2 Feature weight extraction

The different features which we can extract are descibed below:

Word frequency it is a feature used for creating summary, term frequency of a word is calculated by using a fact how commonly a word appears in a document. Inverse document frequency (IDF) depends on the uniqueness or rareness of a word. Word frequency is calculated using a Eq. (1):

$$\mathrm{WF(t)} = \frac{\text{number of times a word appears in a document}}{\text{total number of words in the document}} \tag{1}$$

Title words through this feature we calculate the probability of a sentence having a title word. It is determined by the Eq. (2):

$$\text{probability of title word} = \frac{\text{count of title words in a sentence}}{\text{sentence length}} \tag{2}$$

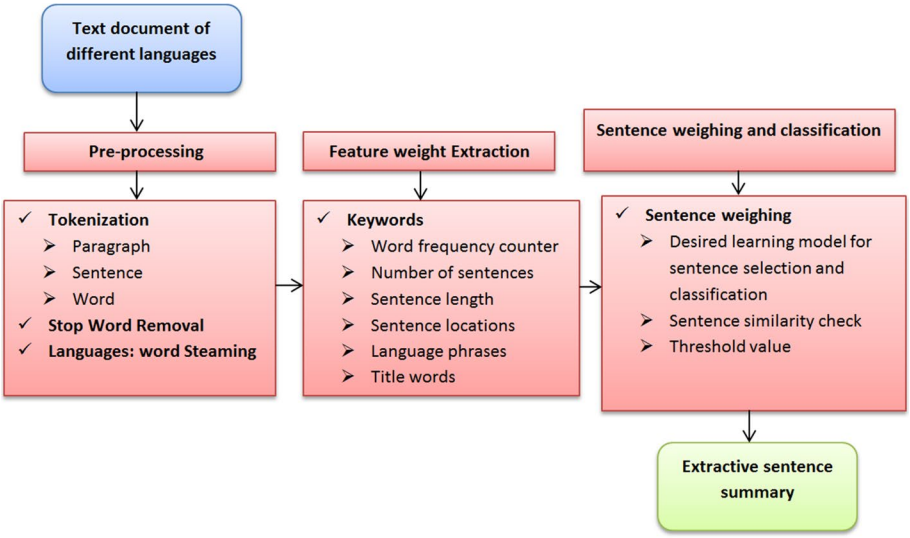


Fig. 3 Framework for text summarization (Bashir et al. 2017)

Sentence length this feature is used to remove the short sentences from the document which are inappropriate in a summay. For example subtitles and datelines in news articles. The sentence length is calculated by the Eq. (3):

$$\text{probability Sentence length} = \frac{\text{word count in the sentence}}{\text{word count in the longest sentence}} \tag{3}$$

Sentence locations it is usually seen that the important sentences fall either in the beginning of the paragraph or in the end, while the redundant sentences lie within the paragraph. The senetence locations is calculated by using Eq. (4):

$$p(S_i, x) = \begin{Bmatrix} \frac{x-(2*y)+1}{x-1}, & y < \frac{x+1}{2} \\ \frac{(2*y)-x-1}{x-1}, & y > \frac{x+1}{2} \\ 0.1, & y = \frac{x+1}{2} \end{Bmatrix} \tag{4}$$

$p(S_i, x)$ gives the probability of sentence based on its location. S_i is the sentence location whereas x is the total count of the sentences in a paragraph.

3.1.3 Sentence weighing and classification

Classification using AI techniques Using the above mentioned features the weight of each sentence is calculated using any disired machine or deep learning model based upon the different laguage text. With the adnacment of artificial intelligent approcahes, the relevent machine or deep learning learning models are being utilized for generating extractive summaries. These techniques use appropriate features which are chosen to create a training model. The trained model is then used for classification to determine whether to include a sentence in summary or not. The techniques that can eb used are CNN, ANN, Fuzzy system (Sudha and Latha 2020).

Sentence Similarity and threshold value In sentence similarity, we eliminate the duplicate sentences using sentence similarity module. Few similarity measures are widely accepted to estimate the degree of similarity between a set of different laguage text files. For example, doc_A and doc_B be the two feature vectors of a set of text documents. The similarity between the documents can be computed using the Eq. (5):

$$Distance\left(doc_A doc_B\right) = \left[\left(doc_A - doc_B\right).\left(doc_A - doc_B\right)\right]^{1/2} \tag{5}$$

Finally, threshold value module is used on the basis of which sentences are picked from the main document to generate the summary. The above three steps are the basic steps involved in the text summarization. These three steps and their corresponding sub-steps define the framework of any text summarization model.

3.1.4 Extractive summary

Summary evaluation is a main feature for text summarization and finally the desired summary is extracted using the assessment task (Mohamed and Oussalah 2019). Usually, intrinsic or extrinsic measures are used to evaluate the summaries. Intrinsic methods measure the quality of summary via human valuation and same is measured in through a task-based performance measure such the information retrieval-oriented task.

3.2 Benefits of automatic text summarization

Several software's are developed that work without human intervention for generating summaries. Following benefits of automatic text summarization are listed below Khan and Naomie (2014):

- Reading or scanning time becomes less.
- While studying any report, outlines make the understanding procedure effortless.
- Summarization enhances the acceptability of ordering.
- Summaries generated algorithmically are less biased compared to human generated summaries.
- Custom-made summaries are more effective in question/answering systems as they deliver personalized data.
- Utilizing programmed or Summarization frameworks allow business theoretical administrations to form the number of content archives they can process.
- Summaries can be generated in any language, as the algorithms can work in any language.
- Summarizing the text leads to increase in productivity and it does not miss any facts.
- More information can be fitted in the small area.

3.3 Challenges in text summarization

Following are the challenges in automatic text summarization:

- One of the factors that makes summarization job relatively challenging is the complication of human language and the way general public express themselves, mainly in transcribed texts. The length of sentence increases as humans use adjectives, adverbs,

appositions etc. to describe their topic. While these can offer valued extra information, they are hardly the key points of the information that we would take in the summary (Eduard and Chin-Yew 1998).

- Moreover, people generally introduce the topic and then try to discuss it using synonyms or pronouns later in the text. Understanding which pronoun is replacement of which earlier presented term is identified in texts as "anaphora problem". Likewise, researcher's face the contrary problem (when unclear words and descriptions discussing a specific term are used in the text former to hosting the term itself) and this is identified as "cataphora problem" (Prasad et al. 2015).
- Furthermore, the source may not contain the text always, such as, a sports event on film or tables presenting trade and industry data, existing tools cannot summarize non-textual content (Lagrini et al. 2017).
- Another challenge is summary evaluation. If you are to believe that the summary is really a trustworthy standby for the source, you must be self-assured that it reflects exactly what is related to the main source. Therefore, approaches for generating and evaluating summaries must accompaniment each other (Saggion and Poibeau 2013).

3.4 Applications of automatic text summarization

Following are the application area of automatic text summarization. Text summarization is turning out to be a good tool in various fields, some of which are mentioned below (Hassel and Dalianis 2012).

- *Media monitoring* Massive amount of data content is available and it is difficult to consume such huge amount of data by the audience, which in a way is a problem. But this problem can be eliminated by Automatic summarization which condense the endless flow of data into reduced but meaningful information (Saggion and Poibeau 2013).
- *Search marketing and SEO* SEO is search engine optimization which is used to increase the visibility of your online content and increase the visitors on your website. So, it required to know what the other competitors are talking about in their content. Multi-document summarization proves to be an influential way to rapidly investigate tons of search results, know shared themes and glance at the key facts.
- *Internal document workflow* Big enterprises are continuously generating in-house information, which commonly gets put in storage and forgotten in records as unstructured information. These enterprises adopt methods that allow re-consume previously prevailing information. Summarization let experts to rapidly comprehend entirety the company has previously prepared in a given theme, and quickly gather information that include diverse viewpoints (Haiduc et al. 2010).
- *Social media marketing* Businesses creating digital content, such as, electronic-books, whitepapers, blogs, might implement summarization to shrink this content and prepare it for sharable on social media sites like Twitter or Facebook.
- *Email overload* Enterprises like Slack were established to save us from relentless emailing. Summarization take us through the main content inside an e-mail and we easily and quickly scan emails.
- *Electronic-learning and tutorials* Lots of teachers use case studies and newscast to structure their course. Summarization assist tutors in swiftly updating their study material by generating shortened information on the topics.

- *Science and Research and Development* Research papers usually consist of a man-made summary that is a condensed form of the paper. It becomes prodigious to declaim all abstracts. Summarization systems cluster papers and shortens the abstracts.

3.5 Sources of datasets in automatic text summarization

The dataset is needed for evaluation of summarization task and compare it with other summarization system to check the applicability of best one (Rodeghero et al. 2014). Several workshops are hosted by Text Analysis Conference (TAC) that gives tracks of various areas of NLP. The **TAC 2014** benchmark contain 20 topics and each have one reference for text and various articles. These are Elsevier published articles in biomedical domain. Other dataset is **CL-Sci Summ** that consist of 30 topics and train, development and test data as three subsets along with one reference paper and articles cites for each topic. Bird have initiated the **ACL Anthology Network** abbreviated as ANN that is a community of people interested in solving the problems related to NLP (Eddy et al. 2013). It contains comprehensive manually curated networked database of citations, summaries and collaborations in the field of Computational Linguistics and ACL published papers. Other is **Microsoft dataset** that is taken from Academic search of Microsoft and consist of information about sentences in the articles abstract, authors, attached paper, citation sentences and place of publication. The **cmp-lg dataset** comprised of 183 documents marked up in XML format and used as resource of extraction and summarization of retrieval information (Sarwadnya and Sonawane 2018). The **PLOS medicine dataset** contains 50 scientific articles each having goof standard summary associated with it. As compared to abstract of article the wider perspective is achieve through summary. The news related datasets are GIGAWORD, CNN daily mail, New York times, Opinion, DUC dataset and many more. The **GIGAWORD dataset** consist of ten million documents of original English Fifth Edition contain articles and their headlines. **CNN daily mail dataset** contains long news articles average of 800 words of both articles and summaries of those articles. Some articles also have summaries of multi-line. The DUC dataset stands for Document understanding conference dataset consist of 500 news articles and their summaries covered at 75 bytes. In this the human authors write the summaries and for single article there is more than one summary. The main limitation of these dataset is that it can't be used in training models with large count of parameters. Other news dataset is New York times contain articles published between 1996 and 2007 and mainly used in extractive systems. The dataset named NEWSROOM is one of the most widely introduced large scale dataset for text summarization. It contains diverse summaries consist of abstractive and extractive strategies. Other datasets are also available such as Google dataset, Webis-TLDR-17, X-Sum, CAST and Ziff-Davis dataset.

4 Reported work

The reported work presented in this section aims to discuss the different text summarizations methods which are widely used by the researchers for automatic text summarization of news articles, summaries, transcripts and text documents. The Fig. 4, shows the text summarization methods.

4.1 Symbolic and theoretical

Currently, the automatic discourse analysis has become an important topic of research due to its use in development of several applications such as information extraction, automatic translation and automatic summarization (Hammad et al. 2016). One of the most employed theories is Rhetorical Structure Theory (RST) but in Spanish very few works has been done by researchers. In 2012, first system assigning rhetorical and nuclearity relations were presented by Cunha et al. (2012), for intra-sentence discourse segments in Spanish texts. The RST Spanish Treebank learning corpus along with manually-annotated specialized texts was analysed by them for building syntactic and lexical patterns marking rhetorical relations. For implementation purpose the DiSeg stands for discourse segment was used with patterns list. Further, the result was compared in terms of precision and recall that shows improved results. Furthermore, RST based automatic summarization technique for Arabic texts were presented by Maaloul et al. (2010). Firstly, the study was done that specify the empirical observations, a set of rhetorical and relations frames. After that the proposed method was applied on texts using ARST Resume system architecture in which the linguistic knowledge was used along with Rhetorical Structure theory. Their complete method depends on three pillars, the first one is location of rhetorical relations between minimal units of the text for which the rhetorical rules was applied (Kamimura and Murphy 2013). The second one is RST-tree simplification and representation that represents the entry of texts in hierarchical form forwarded with senses selection for final summary as third pillar.

4.2 Statistical based text summarization

Vijay et al. proposed a method for Hindi text summarization. They used statistical and linguistic features of the text to generate the summary. The whole process was done in three

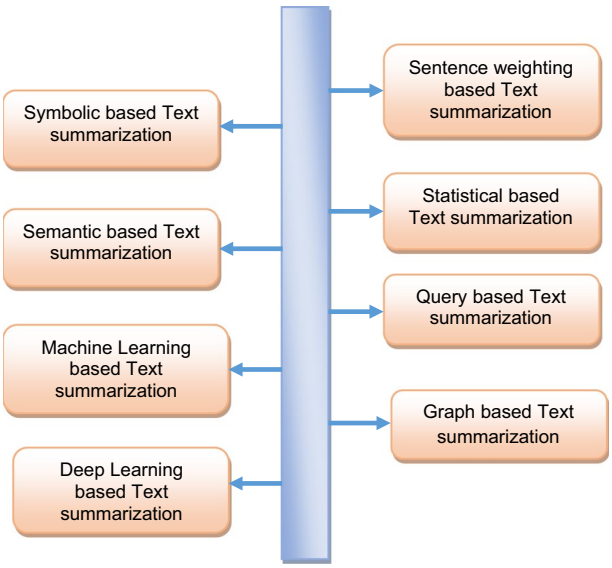


Fig. 4 Mind map of review text summarization

steps pre-processing, processing and post-processing. They used 24,253 News articles as dataset and evaluated generated summary on various parameters. Experimental result shows they obtained recall value as 70 and precision as 62. Afterward, the (Malamos et al. 2005) presented a scheme proficient of creating summaries of Greek newspaper articles, based on statistical methods. Their methodology is divided into three steps which included stemming algorithm, first step, which is capable of extracting stem words and separating the nouns from main document. This leads to domain creation. The second step involved the semantic classification of the target document, known as document classification algorithm. The last step of the algorithm, known as document extraction, generates the summary in a statistic and domain-oriented approach. The simulation result indicates the result in terms of accuracy in comparison to human extracted summaries. For summary size of 30% the proposed method takes 38 s whereas for summary size of 10% it takes 28 s.

Chinese language is an interesting area of research. This language contains more than 50,000 characters and each character is symbolic in nature (Movshovitz-Attias and Cohen 2013). Yu et al. in their research paper described an approach which gathers the original news from an on-line source and generate the summary sentences from them automatically. The basic idea used to extract the sentences from original news used two features structural features and statistical information. They used intrinsic evaluation method to evaluate the generated summaries. To carry out the tests they employed two systems, ATSS-SI (Automatic text summarization system build using both statistical information and structural information) and ATSSS-SS (Automatic text summarization system build using only statistical information). From simulation results they stated ATSS-SI is better than ATSS-SS. The results obtained from ATSS-SI system for summary rate of 10% and 20%, precision = 0.77, recall = 0.78 and precision = 0.74, recall = 0.76 respectively. Further, (Berenjkoob et al. 2009) have presented an algorithm for stemming words and elimination of common words from Persian text summarization. The novelty in their work is their work is free from occurrence of irregular plural words. The database is used from Dehkhoda dictionary. The simulation results showed that the proposed algorithm has high precision and summarize a text by precision of 70% and recall numbers of 3. Furthermore, (Humayoun and Hwanjo 2016) have presented a paper in which they facilitated growth and valuation of single document summarization systems and generated a benchmark corpus. They generated two types of the same corpus. In the type one, they separated words by space. In type two, they manually tagged proper word boundaries. Further they applied normalization, part-of-speech tagging, morphological analysis, lemmatization, and stemming for the articles and their summaries in both types. The evaluation results shown by statistical analysis states that space segmented method showed 64.5% summary coverage whereas properly segmented method showed 65.2% summary coverage.

4.3 Machine learning based text summarization

Machine learning is a way to eliminate the tedious task of creating summaries from human's task list (Fowkes et al. 2017). Machine learning algorithms can be designed and trained to understand text documents and find out the sections that hold the main facts and data former to summary generation. Abstractive and extractive text summarization approaches are examples of machine learning algorithms that help in text summarization. A unique method was presented by Gulati and Sawarkar (2017) that was based on extractive method for multiple Hindi text document summarization. They used fuzzy logic to improve the system performance. They used eleven important features to train

the classifier, some of them includes English–Hindi common words, cue phrases, URLs etc. to generate the summary. The generated summary was found very close to the human generated summaries. The evaluated the system on the basis of three parameters, precision, f-score, recall and they achieved the average precision of about 73% over multiple Hindi documents. Then (Das et al. 2010) have also presented a novel approach for Bengali text summarization based on opinion system using Bengali News corpus. The summary is generated by identifying and aggregating sentiment information from each document for which they used topic-sentiment model. Topic-sentiment model used k-means clustering for sentiment aggregation and Document level Theme Relational Graph representation (DLTRGR). The DLTRGR is actually utilized for sentence selection by standard page rank algorithms used in Information Retrieval (IR) for summary creation. Later on, (Hu et al. 2004) have presented a distinct summarization technique to create single-document summary with complete topic coverage and minimum redundancy. It firstly applied the semantic-class-based vector representations of different types of linguistic units in a document by means of HowNet (an existing ontology), which improved the presentation value of customary term-based vector space model to some extent. Then, they adopted K-means clustering algorithm along with a unique clustering analysis algorithm, where they captured various hidden topic sections in a paper flexibly. Lastly selected characteristic sentences from each segment to generate the final summary. So as to estimate the efficiency of the suggested summarization method, they used a unique metric which is recognized as representation entropy, for redundancy estimation of generated summary. Experimental results obtained shows that the suggested method outperforms the existing basic methods when dealing with varied types of Chinese documents. Breem and Baraka (2017), have also proposed an automatic text summarization approach for multiple documents of Arabic. In proposed work, they have used MapReduce and GA parallel programming model that gives accuracy, speed and scalability in summary generation. Along with it, they have work on elimination of redundancy in sentences that increases the readability and cohesion factors between summaries. This help in getting acceptable value of recall scores and precision that shows a proposed approach successfully identify the most important sentences. Further, (Bois et al. 2014), have described the porting of the English language REZIME (a medical document of single document summarizer) text summarizer to the French language. In this the summaries are created by extraction of key sentences from the original document using ML techniques on lexical, syntactic and statistical features computed on the basis of specialized language resources. The author have presented the architecture of summarizer along with that described the steps needed for adapting the REZIME system to French language.

4.4 Deep learning-based approaches

Baotian et al. (2015) constructed the Chinese short text summarization dataset and used the RNN-methods (recurrent neural network) for summary generation and achieved promising results. They also concluded that performance improves when the input is character based instead of word-based input. They compared the result using ROUGE metric for RNN without context and RNN with context. Further, Sabuna et al. has also presented a scheme which used combination of sentence counting and decision tree method for creating summary. The decision tree algorithm is used for selecting sentences in summary extraction from main content. For training data, they used 50 news texts which were used to produce the rules for decision tree. The used f-measure for result evaluation, the maximum

f-score obtained is 0.80 and the average is 0.58. Again, deep a deep learning approach was used. Bashir et al. (2017) have presented a model for Husa language text summarization. They used naïve bayes model for feature extraction. They used five features in their model for summary generation. They obtained summary size of 30% compression for which they achieved F-score of 78.10%. Furthermore, (Qassem et al. 2019), have improved the performance of Arabic text summarization using fuzzy logic and noun extraction method. For evaluation purpose, they have used EASC corpus and results indicates an improvement compared to existing systems. Parida and Motlicek (2019), have built an abstract text summarizer for German text using transformer model. In their work, they have proposed an iterative data augmentation approach using synthetic data along with real German summarization data. The data set used was Common Crawl German that cover various domains that shows an effectiveness for low resource condition and it is also helpful in their multilingual scenario. Various other issues were also addressed by Straka et al. (2018), and summarization dataset based Czech news was presented by them. The dataset consists of millions of documents containing headline, full text and sentence long abstract. The dataset was evaluated using a language-agnostic variant of ROUGE using Transformer neural network architecture a.

Gulati and Sawarkar (2017) have presented a unique method based on extractive method for multiple Hindi text document summarization. They used fuzzy logic to improve the system performance. They used eleven important features to train the classifier, some of them includes English–Hindi common words, cue phrases, URLs etc. to generate the summary. The generated summary was found very close to the human generated summaries. The evaluated the system on the basis of three parameters, precision, f-score, recall and they achieved the average precision of about 73% over multiple Hindi documents. Further, (Das et al. 2010) have presented a novel approach for Bengali text summarization based on opinion system using Bengali News corpus. The summary is generated by identifying and aggregating sentiment information from each document for which they used topic-sentiment model. Topic-sentiment model used k-means clustering for sentiment aggregation and Document level Theme Relational Graph representation (DLTRGR). The DLTRGR is actually utilized for sentence selection by standard page rank algorithms used in Information Retrieval (IR) for summary creation. Along with it, (Hu et al. 2004) have presented a distinct summarization technique to create single-document summary with complete topic coverage and minimum redundancy. It firstly applied the vector representation of different linguistic units in a document on the basis of semantic class through HowNet. This improves the presentation value of vector space model based customary term to some extent. Then, they adopted K-means clustering algorithm along with a unique clustering analysis algorithm, where they captured various hidden topic sections in a paper flexibly. Lastly selected characteristic sentences from each segment to generate the final summary. So as to estimate the efficiency of the suggested summarization method, they used a unique metric which is recognized as representation entropy, for redundancy estimation of generated summary. Experimental results obtained shows that the suggested method outperforms the existing basic methods when dealing with varied types of Chinese documents.

4.5 Graph and network-based text summarization

Raj and Haroon (2016) have presented an extractive approach for Malayalam text summarization-based graph reduction approach. They used Prim's algorithm for reduction. The key advantage of the proposed system is reduction in redundant information. The database

is collected from online Malayalam news sources. They evaluated results for three parameters, precision, recall, f-score for different datasets. The highest score obtained for dataset1 giving precision of 0.720, recall of 0.818 and f-score of 0.621. Again, (Haroon 2015) have proposed graph theoretic approach for Malayalam text summarization. The result is evaluated for three parameters, precision, recall and f-score for different datasets and the highest value obtained are 0.852, 0.8910 and 0.8018 respectively. The database is collected from Malayalam articles and olam datasets.

Syed and Shanmugasundaram (2017) have also used Graph based approach for Tamil text summarization. They used the page ranking approach and its variants. They compared the generated results with graphical and non-graphical methods. From the results it is evident that directed backward approach performs better than directed forward and undirected schemes. They evaluated the results for effectiveness (E), precision (P), recall value (R). For 10% compression E, P, R holds 0.898, 0.875, 0.727 value and for 30% compression value E, P, R holds 0.794, 0.788, 0.657 respectively. And (Sarwadnya and Sonawane 2018) have used extractive approach for Marathi text summarization. They used ROUGE L for evaluation of generated summary. The experimental result indicates that they have reached quite high f-measures for each approach. TextRank when used along with thematic similarity gives best precision, while TextRank with positional distribution gives best recall. In (Umadevi et al. 2018) have introduced the programmed content calculation synopsis based on extraction approach. In this from first content the coordinated weighted chart is developed the calculation position was used for registration of most critical sentences present in the content and also highlighted in the weighted diagram. The motive of their work was to take news articles and its short summary by considering that no significant information should be left in it. Further, (Mao et al. 2019), have use three methods for extraction of single document summary by combination as supervised and unsupervised learning approach. The motive was to measure the importance of sentences by combination of statistical features of sentences. For scoring the sentences separately a graph and supervised model was utilized by them in first method further final score of sentences was obtained using linear combination of scores. Then graph model was utilized as independent feature of supervised model in second method for evaluation of sentences importance. Then in third method supervised model was utilized for scoring the importance of sentences then sentences are scored using biased graph model. For evaluation purpose a ROUGE method was applied on two datasets named DUC2001 and DUC2022 that shows good results in terms of accuracy using proposed methods as compared to supervised and unsupervised learning.

Afterward, (Bahloul et al. 2019), have presented an A* summarizer unsupervised hybrid approach based automatic system for Arabic single-document summarization. In this the hybridization of graph, cluster and statistical based approach is done. They have divided the text into sub topics after that selected the most relevant sentences from relevant sub topics then A* algorithm was applied for selection that was executed on a graph representing different lexical semantic relationships between sentences. The Essex Arabic summaries corpus was used by them for simulation and compared it with merged model graphs, automatic summarization engineering and n-gram graph powered via regression that shows better results by proposed approach as compared to other.

4.6 Query based text summarization

When there is dealing of heterogeneous documents by users and also wants to compile the important information from it then the use of multi-document summarization proves

to be helpful. In this concern a sentence extraction and clustering based automatic multi-document summarization approach was proposed by Sahoo et al. (2016). In this, on the basis of query provided by user a relevant document is extracted from corpus of document taken from different domains. In the end of complete tasks, a highest weighted top ranked sentence is extracted from each cluster and given to user. This has been found that there are mainly two classification of summarization approach one is extractive and other is abstractive. There is need to understand the original text in case of abstraction summarization after that semantically related summary is generated in it. In this case there is need of understanding a complex natural language processing task. In study done by Mohan et al. (2016), number of works are highlighted in which ontology which is one of approach of abstractive summary was used for abstractive text summarization.

In case of graph-based summarization an important subset of sentences is extracted from a corpus of textual data. In (Baralis et al. 2013) have proposed a new general-purpose graph based summarized and named it GRAPHSUM. In this the association rules were exploited after being discovered for representing the correlations among various terms which used to be neglected by other researchers. The graphs nodes represent a combination of two or more terms are ranked in the basis of PageRank strategy which discriminates the correlation between positive and negative term. Afterward the process of sentence selected was derived using produced node ranking. For analysis purpose, real-life and benchmark documents was used that is further compared with various traditional summarizers. In 2009, (Hendrickx et al. 2009) have presented a MEAD system based automatic multi-document summarization system for Dutch language. For this work, they have focused on redundancy detection as an important part of multi-document summarization. They have also introduced the semantic overlap detection tool that gives better results than other tested methods. Jeong (2013), have also developed a convenient interface that give summary generation, keyword extraction and search engine to users. Their proposed method of summarization was applied to English and Korean news articles and evaluated it via performing various experiments on single and multiple news article user-receptiveness tests along with test collections. The outcome of their proposed approach helps users in more efficiently reading the news articles through various mobile devices.

4.7 Semantic based Text summarization

Dalal and Malik (2017) proposed a system for Hindi text summarization. They used data clustering approach, semantic graph and particle swarm optimization (PSO), in their system for summary generation. They extracted the subject–object–verb (SOV) triples from main text and used them to build the semantic graph of the document and then clustered them into summary and non-summary groups and then classifier generates the main summary which was trained by PSO algorithm. They evaluated the system performance on the recall, precision, F1 score and G score values and obtained 60, 42.86, 50.01, and 50.71 respectively. Further, (Kabeer and Idicula 2014) have also presented a system involving semantic graph approach and statistical sentence scoring scheme for Malayalam text summarization. They evaluated result for both the schemes and evaluated them for precision, recall and f-score value. The results show both the schemes has given comparable results. The statistical method has given precision value as 0.637, recall value as 0.65 and f-score as 0. 533.The semantic method has given precision value as 0.466, recall value as 0.667 and f-score as 0.400. In 2010, (Ozsoy et al. 2010) have found from their survey that Latent Semantic Analysis (LSA) is one of the commonly used methods from all existing one. By

considering it, the explanation of various summarization algorithms based on LSA was explained by them and they have also proposed two new summarization algorithms based on LSA. For evaluation of proposed algorithms, a Turkish document was utilized and their ROUGE-L scores were obtained for comparing their performances that shows best scores obtained using their proposed algorithm. Furthermore, (Mehrnoush et al. 2009) presented an algorithm PARSUMIST for single-document and multi-document summarization using graphs and lexical chains. The algorithm is a combination and improved version of statistical, semantic and heuristic methods. In their experimental results they achieved the precision and recall both about 65%.

For grouping of various Arabic documents into several clusters, hybrid of hierarchical and partitioning based clustering method was proposed by Fejer and Omar (2014). Further from each cluster an important key phrase is extracted using key phrase extraction module that helps in identification of important and similar sentences. They have used ROUGE matrix along with Essex Arabic summaries dataset for evaluation of proposed approach. The evaluation results show that the use of proposed approach gives the accuracy of 80% and 62% for single and multi-document summarization respectively.

4.8 Sentence weighting-based text summarization approach

Gupta and Singh (2012) have presented an extractive approach for Punjabi text summarization which mainly focused on Pre-processing phase and processing phase of Punjabi Text during summarization. Numerous sub phases of pre-processing are: Punjabi words boundary identification, stop words elimination, noun stemming, identification of mutual English Punjabi noun words, identification of proper nouns, identification of sentence boundary, and identification of Punjabi language Cue phrase in a sentence. In processing phase, they identified the weight of each sentence using feature-weight equation. They used Stories News documents as database. They evaluated result for f-score feature and obtained 89.32% f-score for stories database at 30% compression ratio and 95.32% f-score for news documents at 30% compression ratio. Again (Gupta and Singh 2012) have developed an extractive Punjabi summarizer for single document multi news. They are the first to develop summary creation system for Punjabi language and anybody can access this system through an online link: http://pts.learnpunjabi.org/. Also, they are the first to develop many linguistic resources for Punjabi language as part of their project like Punjabi noun morph, Punjabi stemmer and Punjabi named entity recognition, Punjabi keywords identification, normalization of Punjabi nouns etc. Pre-processing phase and processing phase of Punjabi Text during summarization. Numerous sub phases of pre-processing are: Punjabi words boundary identification, stop words elimination, noun stemming, identification of mutual English Punjabi noun words, identification of proper nouns, identification of sentence boundary, and identification of Punjabi language Cue phrase in a sentence. In processing phase, they identified the weight of each sentence using feature-weight equation. The used the extrinsic and extrinsic summary evaluation system for generated summary and obtained, for multi-news single document: f-score = 95.32%, cosine similarity = 96%, question answering task with accuracy = 81.38% (at 30% compression ratio).

Uddin and Khan (2007) have presented an extractive approach for Bangla text summarization. Basically, they did survey of various methods and finally implemented sentence ranking method to generate summary. They considered two parameters to evaluate the summary, summary size and its information. The summary size ranged from 20 to 60% and ranked it on the scale of 0–10 and reported the average highest score of 8.4 at

40% compression ratio. Then (Azmi and Al-Thanyyan 2012) have also presented a method for Arabic text summarization, without using machine learning. They performed the text summarization by using two step algorithms: first step involves the generation of primary summary using Rhetorical Structure Theory (RST); second step involves the assignment of score to each sentence used in the primary summary. The user can limit the size of the final summary. Allotted scores will help in the generation of the final summary. Final summary is generated by selecting the high score sentences with an objective of maximizing the overall score of the summary whose size should not go beyond the user defined limit. To evaluate the generated summaries, they used ROUGE and compared the results with human made (professional editorial persons) summaries. As compare to the Dual Classification method which gives F-measure 0.60 with 30% summary size, their proposed algorithm without any learning process gives F-measure 0.67 with 30% summary size which is appreciable and adds uniqueness.

Kutlu et al. (2010) proposed a standard method for Turkish text summarization by ranking sentences according to their scores. Along with it, a surface level features were used for sentence scores generation and creation of summaries by pulling out the sentences of high rank from the main documents. The sentences were extracted from the main content through title similarity (TS), term frequency (TF), Sentence position (SP) and key phrase (KP) features with less redundancy. For evaluated the comparison was done by generated summary with human-made summaries of Turkish datasets.

They are the first as researchers to familiarize the utilization of KP as a surface-level feature in text summarization and also showed the usefulness of the centrality feature in creating summaries. They acquired 0.561 and 0.368 ROUGE-L scores for both datasets, jounals and newspapers. Cigir et al. (2009) have presented a general Turkish text summarization system utilizing sentences extraction method. They used surface-level document features for instance TF, TS, KP, SP, and centrality of the sentence to find out importance of the sentence. They combined the features and also determined the weights of each feature. They obtained most apt grouping of feature weights by using a training corpus and used this weighing function to extract sentences from main document and then generate the summary using these sentences. The simulation result shows 0.54 recall value and 0.809 precision values with ROUGE evaluation. Then, (Burney et al. 2012) have presented an Urdu text summarization. They used Add-in "Auto Summarizer for Urdu language", especially aimed to create summary of news, informative articles such as sports commentary, economical items, and scientific writings. Mainly their work focused on elimination of stop words. They used sentence weight algorithm to do so. The evaluated result shows that the generated summary was 64% accurate as checked by the human verifiers.

Periantu and Djoko (2017) have presented a scheme which used combination of sentence counting and decision tree method for creating summary. The decision tree algorithm is used for selecting sentences in summary extraction from main content. For training data, they used 50 news texts which were used to produce the rules for decision tree. The used f-measure for result evaluation, the maximum f- score obtained is 0.80 and the average is 0.58. Kutlu et al. (2009) have also proposed a generic text summarization method that generates summaries of Turkish texts by ranking sentences according to their scores. The user surface level features were used for calculating the sentence scores and highest ranked sentences were extracted from original documents to create summaries. For extraction of sentences that form a summary with an extensive coverage of main content of the text from title similarity (TS), term frequency (TF), Sentence position (SP) and key phrase (KP) features with less redundancy. For evaluated the comparison was done by generated summary with human-made summaries of Turkish datasets.

4.9 Different methods-based Text summarization approach

Geetha (2015) presented an approach for Kannada text summarization using Singular Value Decomposition (SVD). SVD is used to find the dimensions of sentence vectors which are orthogonal, both principal and mutually. This scheme assures about relevance of generated summary with original text and also non-redundancy. The system is evaluated for accuracy and precision and scored the 94% and 80% respectively. Then (Fachrurrozi et al. 2013) used the extractive approach for summary generation. The created summaries were compared to the man-made created reference summaries. They considered the counts nouns and verbs as the important parameter for the summary generation. The system also involved statistical approach. They evaluated the generated summary for three parameters, precision, recall and f-measure. Ranking from 1 to 100. Based on the result obtained the proposed was capable of producing summary with f-measure of 78% at the compression rate of 30%. They calculated 83.3% as the average value of quality of generated summary.

Belkebir and Guessoum (2015) have presented a machine learning approach, which used AdaBoost algorithm for Arabic text summarization. They used F1-Score measure for evaluation of the results. In their simulation results they carried out the comparison between three approaches, namely, MLP, AdaBoost, j48. F1-sccore predicts AdaBoost is better than MLP and j48 as the AdaBoost gives the highest score of 66.60% than MLP scores 66.40% and j48 gives the least score of 63.20%. After that, (Jayashree and Murthy 2011) presented two approaches in their work for text summarization, Naïve bayseian and Bag of Words (BOW) and results proved BOW is significantly better than Naïve Bayseian approach giving high precision value but low recall value. They obtained overall precision of 89.05% using BOW model. Hidayat et al. (2015) have presented an approach for text summarization using Latent Drichlet al.location to advance accurateness in document clustering. The test involved 398 datasets from public blog article. They used some steps of clustering, namely, preprocessing, automatic data shrink using feature method and using LDA, word weighting and clustering algorithm. They evaluated result for two different techniques with LDA they obtained 72% accuracy whereas with traditional k-means method they obtained 66% of accuracy.

Pontes et al. (2018) have introduced the use of two compression approaches along with chunks for generation of more informative linguistic summaries. For testing the proposed approach, a MultiLing 20,122 dataset was used that shown an improvement from traditional approaches. Liu and Wang (2017) have addressed some of the problem of Korean text summarization (KTS) and presents a flexible multi-plugin framework. Within the framework, they designed a novel KTS algorithm based on key phrase extraction. Supported by the pluggable components of word stemming and part of speech tagging, the key-phrase-extraction-based KTS algorithm can complete text summarization efficiently. The experimental results show that the proposed KTS algorithm with MMR plugin component can achieve the perfect performance in the Korean summarization task. Further, (Lehto and Sjodin 2019), have described a modified TextRank model and investigated the different methods available to use automatic text summarization as a means for summary creation of Swedish news articles. The proposed method focused on intrinsic evaluation methods, in part through content evaluation in the form of measuring referential clarity and non-redundancy, and in part by text quality evaluation measures, in the form of keyword retention and ROUGE evaluation.

Table 1 Comparative analysis of ATS in different Indian language

Author	Summarization languages	Dataset	Technique used	Challenges	Parameters
Dalal and Malik (2017)	Hindi	80 Hindi documents	Data clustering approach, semantic graph and particle swarm optimization (PSO)	The suitability of the approach need to be improved in case of integrated co-reference and anaphora resolution in the pre-processing phase	Recall = 60 Precision = 42.86 F1 score = 50.01 G score = 50.71
Kabeer and Idicula (2014)	Malayalam	25 documents from Malayalam News sources	Semantic graph approach, semantic method	They have only tested the text summarization for Malayalam documents using semantic approaches only. Other approaches need to apply for more improvement	For semantic graph approach: precision = 0.637, Recall = 0.650 F-measure = 0.533 For semantic method precision = 0.466, Recall = 0.667 F-measure = 0.400
Syed et al. (2017)	Tamil	50 Random Tamil documents	Graph based approach	They have not used multi document sets with other language processing tools	For 30% compression ratio: E = 0.794 P = 0.788 R = 0.657
Sarwadnya and Sonawane (2018)	Marathi	634 News articles	Graph based Approach	Need to extend the scope by using other summarization such as abstractive summarization	ROUGE 1; TextRank + positional, F-score = 0.8004 TextRank + similarity, Fscore = 0.8684
Raj and Haroon (2016)	Malayalam	Online Malayalam news sources	Minimum spanning tree-based graph reduction approach	They have used basic dataset for evaluation that need to be improved by adding good data dictionary	Precision = .720, Recall = 0.818 F-measure = 0.621
Bahloul et al. (2019)	Arabic	Essex Arabic Summaries	Combination of statistical, cluster-based, and graph-based techniques	They have only used lexical and synonymy repetitions in text segmenter	Recall = 0.4784, Precision = 0.5396, F-measure = 0.4940

Table 1 (continued)

Author	Summarization languages	Dataset	Technique used	Challenges	Parameters
Haroon (2015)	Malayalam	Malayalam articles and olam datasets	PARSUMIST (combination of statistical, semantic and heuristic method)	There are various advanced features that need to be added in the work for more effectiveness of summaries	Precision and recall both = 65%
Vijay et al. (2017)	Hindi	In-short online news for Hindi text summarization	Statistical and linguistic features	No as such comparison is given with other models that need to be done for more validation of proposed work	Recall = 70 Precision = 62
Humayoun and Hwanjo (2016)	Urdu	Urdu summary corpus using 50 articles from various domains	Statistical analysis of Space segmentation (SS) and proper segmentation (PS)	In existing work only one abstractive summary is considered per article that needs to be increased	SS = 64.5% coverage PS = 65.2% coverage
Gulati and Sawarkar (2017)	Hindi	News articles from online Hindi newspapers	Fuzzy logic classifier and eleven features	They have only implemented a single algorithms on Hindi text	Average precision = 73%
Das et al. (2010)	Bengali	Bengali News corpus	Classifier: K-means clustering features, lexico-syntactic, syntactic, Discourse level	The hierarchal cluster of theme words need to be generated for targeted work	precision = 72.15%, Recall = 67.32% F-measure = 69.65%
Gupta and Singh (2013)	Punjabi	50 News documents	Sentence weighting score	More news documents can be used for validation of proposed work	For multi-news single document: F-score = 95.32% cosine similarity = 96% Q/A task with accuracy = 81.38%

Table 1 (continued)

Author	Summarization languages	Dataset	Technique used	Challenges	Parameters
Uddin and Khan (2007)	Bengali	Bangla Newspaper Text	Sentence ranking	The main limitation is that in Bangla summarizer only few sentences are extracted from given text that is not similar to summary generated by human	Accuracy = 84%
Burney et al. (2012)	Urdu	25 randomly selected documents	sentence weight algorithm	There is need to implement the advanced algorithms for improving the efficiency of Auto-Summarizer	Accuracy = 64%
Geetha (2015)	Kannada	Database is collected from Arts, Commerce, Homeopathy, Literature	Singular Value Decomposition	There is no processing syntax and representation of LSA that result in reduction of performance for multilingual and large document text	Accuracy = 94% Precision = 80%
Jayashree and Murthy (2011)	Kannada	Kannada Wikipedia	Bag of Words (BOW) approach	There is no sufficient class information as a standalone entity	Precision = 89.05%
Baruh et al. (2020)	Assamese	Assamese language corpus	ROUGE Model	The content compaction technique for Assamese text to produce summary by statistical and linguistic features is not easy task	F-measure = 0.633

Table 1 (continued)

Author	Summarization languages	Dataset	Technique used	Challenges	Parameters
Rodrigues et al. (2019)	Konkani	Konkani documents	Sentence extraction technique	Limited features has been used to extract the Konkani language document which resulted not extracting the important sentences of the language	Recall = 0.250 Precision = 0.152 F-Score = 0.189
D'Silva and Sharma (2020)	Konkani	Konkani language corpus	K-means algorithm	The proposed system for text-summarization of Konkani language does not utilize the language dependent domain knowledge such as stop-words, stemming and other training corpus	Recall = 0.358 Precision = 0.354 F1-Score = 0.356
Ranabhat et al. (2019)	Nepali	Nepalese news article	TextRank method	The machine and deep learning model can be applied to train the system efficiently to generate summary based on Nepalese news article	Avg. Summary Score = 5.5.865
Balabantaray et al. (2012)	Odia	Odia news data	Extractive text summarization	The system is trained for only 20 documents having 30 sentences of Odia news text	Sentence score = 0.615
Biswas et al. (2015)	Odia	Online Odia documens	Web Frequency method	The proposed Oriya automatic text summarizer van be further extended by adding abstractive summaries	Precision = 95%
Sakhare and Kumar (2016)	Sanskrit	DUC 2002 data set	Neural Network	The deep learning models can be applied for training the system and to obtain the better outcomes	Rough-1 value = 0.234 Rough-2 value = 0.0816

Table 1 (continued)

Author	Summarization languages	Dataset	Technique used	Challenges	Parameters
Nathani et al. (2020)	Sindhi	Sindhi Dictionary	Minimum description Length (MDL) algorithm	The over stemming and under stemming error can be overcome by applying lemmatization rules	Accuracy = 87%
Ali and Wagan (2017)	Sindhi	Sindhi sentiment text	SVM KNN	The visual tracking can be done in Sindhi sentiment analysis to get proper opinions	Avg. Precision = 89% for SVM and 68% for KNN
NagaPrasad et al. (2015)	Telugu	300 Telugu news articles	SVM	The work can be extended and improved by adding more features and machine learning models in the authorship attribution of Telugu text	F1-Score = 0.82 Accuracy = 0.87
Sudha and Latha (2020)	Talugu	Heterogeneous multi-document datasets	RNN	The more learning models can be applied for prediction and improvement of the work	F-Score = 0.029 Precision = 0.016 Recall = 0.041
Ramanujam and Kaliappan (2016)	Gujarati	Gujarati text corpus	Naive Bayesian	The automatic text summarizer based on Naïve Bayes can be further improved further adding more features and machine learning models	Precision = 85.4% Recall = 83.9% F-measures = 92%T

Table 2 Comparative analysis of ATS in Foreign languages

Author	Summarization languages	Dataset	Technique used	Challenges	Parameter used
Mehrnoush et al. (2009)	Persian	Persian documents	Statistical technique PARSUMIST (Combination of statistical, semantic and heuristic method)	The score of proposed method is very less due to absence of consideration on type of input files like article, stories and newspaper	precision = 70% Recall = 3, Precision and recall both = 65%
Umadevi et al. (2018)	Spanish	Two Columbian presidents talks transcript	Modified page rank algorithm and weighted graph approach	There is issue of best approach for getting applicable course of action in the existing framework	Similarity = 100
Yu et al. (2006)	Chinese	AC news for collection of 30 Chinese news articles	Structural features and statistical information	There is need of better method for generating summaries that are better understood by human beings	ATSS-SI system For 10% summary rate Precision = 0.77; recall = 0.78
Malamos et al. (2005)	Greek	1080 Greek articles taken from newspapers	Statistic based algorithm	There is need of translator for different languages in translation phase for directly discussing various difficult cases	For summary size of 30%, time taken = 38 s for summary size of 10%, time taken = 28 s
Berenjkoob et al. (2009)	Persian	Dehkhoda dictionary	Statistical technique	There are some limitations of streaming algorithms	precision = 70% Recall = 3
Hu et al. (2004)	Chinese	Constructed from Chinese microblogging website Sina Weibo	RNN (recurrent neural networks) with and without context	There is need of large document summarization data set made by natural annotated web resources	RNN with context Character based approach ROUGE-1 = 0.299

Table 2 (continued)

Author	Summarization languages	Dataset	Technique used	Challenges	Parameter used
Bashir et al. (2017)	Hausa	10 Hausa language documents	Features: sentence length, sentence location, title words, word frequency, no. of sentences for a word Classifier: Naïve bayes model	There is need of more semantic features like part of speech	F-score = 78.10%
Qassem et al. (2019)	Arabic	EASC	Fuzzy logic	The existing Arabic preprocessing is not very efficient that creates a need of better one for getting accurate outcomes for Arabic summarization	Average recall = 0.45, Average precision = 0.48, Average F-measure = 0.44
Straka et al. (2018)	Czech	Czech news-based dataset	Transformer neural network	The quality produced by t2t method in case of lower Rouge raw sources in the second case	For t2t method precision = 7.4, recall = 5.9 and f-measure = 6.4
Bois et al. (2014)	French and English	English BioMed central and French Acta Endoscopica	Machine learning techniques, using statistical, syntactic and lexical features	The use of proposed approach will get affected by lack of parallel data and consequently use of varied evaluation collections	Rouge score for English = 11 and French = 10
Breem and Baraka (2017)	Arabic	Reports dataset	Genetic algorithm and MapReduce parallel programming model	There is some limitations of existing approach such as addition of identifying named entity need to be improved	Precision = 0.57, Recall = 0.21, F-measure = 0.31

Table 2 (continued)

Author	Summarization languages	Dataset	Technique used	Challenges	Parameter used
Kopeć (2019)	Polish	Polish Summaries Corpus	Supervised Machine learning models	There is need to test the use of automatic text summarization in complex task	ROUGE = 0.339
Jassem and Pawluczuk (2015)	Polish	Polish Summaries Corpus	Neural Networks	There are some challenges associated with choosing the proper set of features and tuning ML algorithms	Recall = 0.51 Precision = 0.45 F-1 = 0.56
Lee et al. (2020)	Korean	Korean Daum/News dataset	Pre-trained neural network (SBERT)	The proposed method is used for Korean language only that needs to demonstrated on English summarization dataset	Pearson correlation = 0.38 Kendall Kendall rank = 0.22
Hendrickx et al. (2009)	Dutch	Parallel/comparable text articles	Query-based document clusters	There is need of more refined vision that uses word alignments and alignments at parse tree level	Macro average Rouge scores = 0.150
Jeong (2013)	Korean	KORDIC	Query expansion technique	The main challenge of existing work is conversion of standard web pages on smart phone in case of large amount of content on each page	F1-measure = 0.511
Ren and Kang (2018)	Korean	1,000 Internet news items	Rule-Based Method	The extraction performance of MII need to be improved and need to apply the key information for obtaining abstractive summaries	Precision = 0.83, Recall = 0.75, F-measure = 0.788

Table 2 (continued)

Author	Summarization languages	Dataset	Technique used	Challenges	Parameter used
Azmi and Al-Thanyyan (2012)	Arabic	32 documents from various news articles, and Saudi newspapers	RST-based system and a sentence scoring scheme	Very few features are used in existing work	F-score = 0.67
Kutlu et al. (2010)	Turkish	Turkish jounals and newspapers	Surface level features: frequency, key phrase (KP), centrality, title similarity and sentence position	Very less focus is giving in observing the effects of key phrases by changing their scoring features	0.561 and 0.368 ROUGE-L scores for both datasets
Fejer and Omar (2014)	Arabic	Essex Arabic Summaries	Similarity algorithm	They have not done the comparison of their work with other systems that need to be done	Accuracy for single document = 80% and Multi-document summarization = 62%
Periantu and Djoko (2017)	Indonesia Bhasha	50 news texts domains	Sentence scoring and decision tree	There is need of professionals manual and document corpus summary for improving standardization of testing	Average f-score = 0.58
Cigir et al. (2009)	Turkish	Online Turkish Newspaper	Sentence extraction features: TF, TS, KP, SP, and centrality of the sentence	Very less study has been done in the proposed work	Recall = 0.54 precision = 0.809
Cunha et al. (2012)	Spanish	RST Spanish Treebank,	Symbolic Approach	The rules used in existing work need to be optimized	Precision = 81.75, Recall = 81.75
Fachrurrozi et al. (2013)	Indonesia bhasha	Human produced summary and 30 articles as reference	Term Frequency	There is some limitation in the existing system that need to be improve by integration of system with some another method	Summarized Text Understood by 83.3% respondents

Table 2 (continued)

Author	Summarization languages	Dataset	Technique used	Challenges	Parameter used
Belkebir and Guessoum (2015)	Arabic	Real-time data	AdaBoost	There is need of testing the proposed approach on extended database consists of more documents	F1-sccore = 66.60%
Hidayat et al. (2015)	Indonesia bhasha	398 articles from various domains	Latent Dirichlet al.location (LDA)	The proposed system is not much accurate that need to be improve	Highest average accuracy = 72%, at compression rate of 40%
Pontes et al. (2018)	Czech and multi-lingual	MultiLing 2011 dataset	Two compression methods at the sentence and multi-sentence levels	To show the improvements in term of grammar and informativeness there is need of additional human evaluation	F-measure of multi-sentence compression version system = 0.47221
Parida and Motlicek (2019)	German	German wiki data from the SwissText Common Crawl	State-of-the-art "Transformer" model	Very less work has been done on synthetic summarization data and applying transfer learning on text summarization for multilingual low resource dataset	The average length of words generated by model S2 is longer than the model S1
Liu and Wang (2017)	Korean	Korean journals	Korean text summarization (KTS) algorithm	The proposed method causes various disfluencies between sentences in the summary	Recall-oriented understudy for gisting evaluation (ROUGE) = 0.035
Lehto and Sjodin (2019)	Swedish	Swedish news articles	TextRank Model	In existing work less information is used as compared to TextRank that creates a need of taking referential clarity into consideration	Precision = 0.35 for modified TextRank

5 Comparative analysis

Under this section, comparative analysis on automatic text summarization is presented on both Indian and foreign languages. India is a multi-cultural and multi-religious country. It is a site of diverse lingual families (Shah and Desai 2016). India has 22 languages, namely, Hindi, Punjabi, Sanskrit, Tamil, Kannada, Marathi, Telugu, Urdu, Bengali, Oriya, Guajarati, Sindhi, Assamese, Malayalam, Manipuri, Konkani, Nepali, Odia and many more (Baruah et al. 2019). Out of these most of the languages work we have covered. Along with it, various non-Indian languages are considered named Chinese, Rome, Czech, Spanish, French, Polish, German, Turkish, Swedish, Arabic and Hausa.

From the Tables 1 and 2, it can be concluded that significant work has been done in context to text summarization in different languages using various text summarization methods. Based on the study, we have found that most of the work has been done in ATS using machine and deep learning-based classifications approaches and obtained the optimal results in various parameters.

6 Conclusion and future work

This paper is an attempt to provide a brief overview about the research carried out in the field of text summarization in different languages using various text summarization methods. We have also discussed about various applications, approaches, different datasets and challenges of text summarization. The paper presented a collection of good findings with adequate accuracy, precision, and effectiveness, in both Indian languages such as, Hindi, Punjabi, Bengali, Malayalam, Kannada, Tamil, Marathi, Assamese, Konkani, Nepali, Odia, Sanskrit, Sindhi, Telugu and Gujarati along with foreign languages such as Arabic, Chinese, Greek, Persian, Turkish, Spanish, Czeh, Rome, Urdu, Indonesia Bhasha and many more. The shortage of a standard database on various Indian languages is also a major concern. Tabular representation for comparative study of the reported work in order to assist the novice researchers in this field. Through this paper, we also highlighted number of features and classifiers for summarization of the text. This work also help researcher in getting thorough study on various approaches of text summarization.

For future directions, it can be expected to find novel ways for the feature extraction and classifiers for achieving the best results and accuracy for text summarization. The availability of database for particular languages is also a concern and a novel future aspect in this field. Developing hybrid models for single document along with multi-document text summarization is also an area of research and important issue for researchers.

References

Ali M, Wagan AI (2017) Sentiment summerization and analysis of Sindhi text. Int J Adv Comp Sci Appl, pp 296–300.

Azmi AM, Al-Thanyyan S (2012) A text summarizer for Arabic. Computer Speech Lang, pp 260–273.

Bahloul B, Aliane H, Benmohammed M (2019) ArA*summarizer: An Arabic text summarization system based on subtopic segmentation and using an A* algorithm for reduction. Wiley Expert systems, New York, pp 1–16.

Balabantaray RC, Sahoo B, Sahoo DK, Swain M (2012) Odia text summarization using stemmer. Int J Appl Inf Syst (IJAIS), pp 21–24.
Baotian H, Qingcai C, Fangze Z (2015) Lcsts: a large scale Chinese short text summarization dataset. arXiv preprint arXiv, pp 1–6.
Baralis E, Cagliero L, Mahoto N, Fiori A (2013) GRAPHSUM: discovering correlations among multiple terms for graph-based summarization. Inf Sci, pp 96–109.
Baruah N, Sarma S, Borkotokey S (2019) Text summarization in Indian languages: a critical review. In: IEEE second international conference on advanced computational and communication paradigms (ICACCP), pp 1–6
Baruah N, Sarma SK, Borkotokey S (2020) Evaluation of content compaction in Assamese language. Third international conference on computing and network communications (CoCoNet'19), pp 2275–2284.
Bashir M, Rozaimee A, Wan M, Isa W (2017) Automatic Hausa language text summarization based on feature extraction using Naïve Bayes model. World Appl Sci J 35(9):2074–2080
Belkebir R, Guessoum A (2015) A supervised approach to Arabic text summarization using adaboost. In: Springer New contributions in information systems and technologies, pp 227–236
Berenjkoob M, Mehri R, Khosravi H, Nematbakhsh MA (2009) A method for stemming and eliminating common words for Persian text summarization. In: IEEE International conference on natural language processing and knowledge engineering, pp 1–6
Bhatia N, Jaiswal A (2016) Automatic text summarization and it's methods-a review. In: IEEE 6th international conference-cloud system and big data engineering (Confluence), pp 65–72
Biswas S, Acharya S, Dash S (2015) Automatic text summarization for Oriya language. Int J Comp Appl, pp 19–26.
Bois R, Levelling J, Goeuriot L, Jones GJF, Kelly L (2014) Porting a summarizer to the French language. 21ème Traitement Automatique des Langues Naturelles, Marseille, pp 550–555.
Breem SN, Baraka RS (2017) Automatic arabic text summarization for large scale multiple documents using genetic algorithm and MapReduce. In: Palestinian International Conference on Information and Communication Technology, pp 40–45.
Burney A, Sami B, Mahmood N, Abbas Z, Rizwan K (2012) Urdu text summarizer using sentence weight algorithm for word processors. Int J Comp Appl, pp 38–43
Cigir C, Kutlu M, Cicekli I (2009) Generic text summarization for Turkish. In: IEEE 24th International symposium on computer and information sciences, pp 224–229
Cunha ID, Juan ES, Torres-Moreno J-M, Cabre MT, Sierra G (2012) A symbolic approach for automatic detection of nuclearity and rhetorical relations among intra-sentence discourse segments in Spanish. In: International conference on intelligent text processing and computational linguistics. Springer, Heidelberg, pp 462–475.
D'Silva, J, Sharma U (2020) Unsupervised automatic text summarization of Konkani texts using K-means with Elbow method. Int J Eng Res Technol, pp 2380–2384.
Dalal V, Malik L (2017) Data clustering approach for automatic text summarization of Hindi documents using particle swarm optimization and semantic graph. In: International Journal of Soft Computing and Engineering (IJSCE), pp 1–3.
Das A, Bandyopadhyay S (2010) Topic-based Bengali opinion summarization. In: Proceedings of the 23rd international conference on computational linguistics: posters, pp 232–240
Eddy BP, Robinson JA, Kraft NA, Carver JC (2013) Evaluating source code summarization techniques: replication and expansion. In: 21st International conference on program comprehension (ICPC), IEEE, pp 13–22.
Eduard H, Chin-Yew L (1998) Automated text summarization and the SUMMARIST system. workshop on held at Baltimore. Maryland, Association for Computational Linguistics, pp 197–214
Fachrurrozi M, Yusliani N, Yoanita RU (2013) Frequent term-based text summarization for bahasa Indonesia. In: International Cconference on innovations in engineering and technology (ICIET'), pp 30–32
Fejer HN, Omar N (2014) Automatic Arabic text summarization using clustering and keyphrase extraction. In: International conference on information technology and multimedia (ICIMU), pp 293–298.
Florescu C, Jin W (2019) A supervised keyphrase extraction system based on graph representation learning. European conference on information retrieval, pp 197–212.
Fowkes J, Chanthirasegaran P, Ranca R, Allamanis M, Lapata M, Sutton C (2017) Autofolding for source code summarization. IEEE Trans Softw Eng, pp 1095–1109
Geetha JK, (2015) Kannada text summarization using Latent Semantic Analysis. In: IEEE International conference on advances in computing, communications and informatics (ICACCI), pp 1508–1512
Gulati AN, Sawarkar SD (2017) A novel technique for multi-document Hindi text summarization. In: IEEE International conference on Nascent technologies in engineering (ICNTE), pp 1–6

Gupta V, Singh GL (2012) Automatic Punjabi text extractive summarization system. In: Proceedings of COLING: Demonstration Papers, pp 191–198
Gupta V, Singh GL (2013) Automatic text summarization system for Punjabi language. J Emerg Technol Web Intell, pp 257–271
Haiduc S, Aponte J, Marcus A (2010) Supporting program comprehension with source code summarization. In: Proceedings of the 32nd ACM/IEEE international conference on software engineering. ACM, New York, pp 223–226.
Hammad M, Abuljadayel A, Khalaf M (2016) Summarizing services of java packages. Lecture Notes on Software Engineering, pp 129–132.
Haroon RP (2015) An extractive Malayalam document summarization based on graph theoretic approach. In: IEEE Fifth international conference on e-Learning (econf), pp 237–240
Hassel M, Dalianis H (2012) Portable text summarization. In: Applied natural language processing: identification, investigation and resolution, pp 17–32
Hendrickx I, Daelemans W, Marsi E, Krahmer E (2009) Reducing redundancy in multi-document summarization using lexical semantic similarity. In: Proceedings of the 2009 Workshop on language generation and summarisation, ACL-IJCNLP, pp 63–66.
Hidayat EY, Firdausillah F, Hastuti K, Ika ND, Azhari (2015) Automatic text summarization using latent drichlet allocation (LDA) for document clustering. Int J Adv Intell Informatics, pp 132–139
Hu P, Tingting H, Donghong J, Meng W (2004) A study of Chinese text summarization using adaptive clustering of paragraphs. In: IEEE Fourth international conference on computer and information technology, pp 1159–1164
Humayoun M, Hwanjo Y (2016) Analyzing pre-processing settings for Urdu single-document extractive summarization. In: Proceedings of the tenth international conference on language resources and evaluation (LREC), pp. 3686–3693
Jassem K, Pawluczuk L (2015) Automatic summarization of Polish news articles by sentence selection. Federated Conference on Computer Science And Information Systems, pp 1–5.
Jayashree, Murthy KS (2011) An analysis of sentence level text classification for the Kannada language. In: IEEE International conference of soft computing and pattern recognition (SoCPaR), pp. 147–151
Jeong H (2013) Efficient keyword extraction and text summarization for reading articles on a smart phone. Comput Informatics, pp 1001–1016.
Kabeer R, Idicula MS (2014) Text summarization for Malayalam documents-an experience. In: IEEE International conference on data science & engineering (ICDSE), pp 145–150
Kamimura M, Murphy GC (2013) Towards generating human-oriented summaries of unit test cases. In: 21st International conference on program comprehension (ICPC), IEEE, pp 215–218.
Khan A, Naomie S (2014) A review on abstractive summarization methods. J Theor Appl Inf Technol, pp 64–72
Kopeć M (2019) Three-step coreference-based summarizer for Polish news texts. Poznań Studies in Contemporary Linguistics, pp 397–443.
Kutlu M, Cıgır C, Cicekli I (2010) Generic text summarization for Turkish. Comp J, pp 1315–1323
Lagrini S, Redjimi M, Azizi N (2017) Automatic arabic text summarization approaches. Int J Computer Appl, pp 31–37
Lee D, Shin M, Whang T, Cho S, Ko B, Lee D, Kim E, Jo J (2020) Reference and Document Aware Semantic Evaluation Methods for Korean Language Summarization. pp 1–13
Lehto N, Sjodin M (2019) Automatic text summarization of Swedish news articles. Eng Tech 1–12
Liu W, Wang L (2017) Efficient Korean text summarization based on key phrase extraction. In: International conference on machine learning and cybernetics, pp 61–66.
Maaloul MH, keskes I, Belguith LH, Blache P (2010) Automatic summarization of Arabic texts based on RST technique. In: Proceedings of the 12th international conference on enterprise information systems, pp 1–7.
Malamos AG, Ware MGJA (2005) Applying statistic-based algorithms for automated content summarization in Greek language, Jaoua, Ben, pp 1–8
Mao X, Yang H, Huanga S, Liua Y, Li R (2019) Extractive summarization using supervised and unsupervised learning. Expert systems with applications, pp 173–181, 2019.
Mehrnoush S, Tara A, Erfani JM (2009) Parsumist: a Persian text summarizer. In: IEEE International conference on natural language processing and knowledge engineering, pp 1–7.
Mihalcea R, Tarau P (2004) TextRank: bringing order into texts. Empirical methods in natural language processing (EMNLP). Barcelona, Spain, pp 404–411.
Mohamed M, Oussalah M (2019) SRL-ESA-TextSum: a text summarization approach based on semantic role labeling and explicit semantic analysis. Information Processing & Management, pp 1356–1372.

Mohan MJ, Sunitha C, Ganesha A, Jaya A (2016) A study on ontology based abstractivesSummarization. Procedia Computer Science, pp 32–37.
Moratanch N, Chitrakala S (2017) A survey on extractive text summarization. In: 2017 International conference on computer, communication and signal processing (ICCCSP), pp 1–6.
Movshovitz-Attias D, Cohen WW (2013) Natural language models for predicting programming comments. In: Proceedings of the 51st annual meeting of the association for computational linguistics, pp 35–40.
Nagaprasad S, Vijayapal Reddy P, Vinaya Babu A (2015) Authorship Attribution based on Data Compression for Telugu Text. Int J Comput Appl 110(1):1–5
Nallapati R, Zhai F, Zhou B (2017) SummaRuNNer: a recurrent neural network based sequence model for extractive summarization of documents. Thirty-First aaai conference on artificial intelligence (AAAI-17), pp 3075–3081.
Nathani B, Joshi N, Purohit GN (2020) Design and development of unsupervised Stemmer for Sindhi language. In: International Conference on Computational Intelligence and Data Science (ICCIDS), pp. 1920–1927.
Ozsoy MG, Cicekli I, Alpaslan FN (2010) Text summarization of Turkish texts using latent semantic analysis. In: Proceedings of the 23rd International conference on computational linguistics, pp 869–876.
Parida S, Motlicek P (2019) Idiap abstract text summarization system for German text summarization task. SwissTex, pp 1–5.
Parveen D, Mesgar M, Strube M (2016) Generating coherent summaries of scientific articles using coherence patterns. Empirical methods in natural language processing, Texas: Austin, pp 772–783.
Periantu MS, Djoko BS (2017) Summarizing Indonesian text automatically by using sentence scoring and decision tree. In: IEEE 2nd International conferences on information technology, information systems and electrical engineering (ICITISEE), pp 1–6
Pontes EL, Huet S, Torres-Moreno J-M, Linhares AC (2018) Cross-language text summarization using sentence and multi-sentence compression. Natural Language Processing and Information Systems, pp 467–479.
Prasad, SN, Narsimha, VB, Reddy, PV, Babu, AV (2015) Influence of lexical, syntactic and structural features and their combination on Authorship Attribution for Telugu Text. In: International conference on intelligent computing, communication & convergence, pp 58–64.
Qassem LMA, Wanga D, Barada H, Rubaiea AA, Moosaa NA (2019) Automatic Arabic text summarization based on fuzzy logic. In: Proceedings of the 3rd international conference on natural language and speech processing, pp 42–48.
Raj MR, Haroon RP (2016) Malayalam text summarization: minimum spanning tree-based graph reduction approach. In: IEEE 2nd International conference on advances in computing, communication, & automation (ICACCA) (Fall), pp 1–5
Ramanujam, N, Kaliappan, M (2016) An automatic multidocument text summarization approach based on Naive Bayesian classifier using timestamp strategy. Sci World J, pp 1–11
Ranabhat R, Upreti A, Sangpang B, Manandhar S (2019) Salient sentence extraction of Nepali online health news texts. Int J Adv Soc Sci, pp 21–26.
Ren M, Kang S (2018) Korean news text summarizer enriched with major information items. Int J Adv Sci Technol, pp 115–126.
Rodeghero P, McMillan C, McBurney PW, Bosch N, Mello SD (2014) Improving automated source code summarization via an eye-tracking study of programmers. In:Proceedings of the 36th international conference on Software engineering, pp 390–401.
Rodrigues S, Fernandes S, Pai A (2019) Konkani text summarization by sentence extraction. In: 10th International conference on computing, communication and networking technologies (ICCCNT), pp 1–6.
Saggion H, Poibeau T (2013) Automatic text summarization: past, present and future. In: Multi-source, multilingual information extraction and summarization. Springer, Heidelberg, pp 3–21
Sahoo D, Balabantaray R, Phukon M, Saikia S (2016) Aspect based multi-document dummarization. In: International conference on computing, communication and automation (ICCCA2016), pp 873–877.
Sakhare DY, Kumar R (2016) Syntactical knowledge and Sanskrit memamsa principle based hybrid approach for text summarization. Int J Comp Sci Inf Security (IJCSIS), pp 270–275.
Sarwadnya VV, Sonawane SS (2018) Marathi extractive text summarizer using graph based model. In: IEEE Fourth international conference on computing communication control and automation (ICCUBEA), pp 1–6
Shah P, Desai N (2016) A survey of automatic text summarization techniques for Indian and foreign languages. IEEE International conference on electrical, electronics, and optimization techniques (ICEEOT), pp 4598–4601
Shimpikar S, Govilkar S (2017) A survey of text summarization techniques for Indian regional languages. int J Comp Appl, pp. 29–33

Straka M, Mediankin N, Kocmi T, Zabokrtsky Z, Hudecek V, Ha J (2018) SumeCzech: large Czech News-based summarization dataset. In: Proceedings of the eleventh international conference on language resources and evaluation (LREC), pp 3488–3495.

Sudha, DN, Latha YM (2020) Multi-document abstractive text summarization through semantic similarity matrix for Telugu language. Int J Adv Sci Technol, pp 513–521.

Syed SM, Shanmugasundaram H (2017) An investigation on graphical approach for tamil text summary generation. In: IEEE International conference on intelligent computing and control (I2C2), pp 1–5

Uddin MN, Khan SA (2007) A study on text summarization techniques and implement few of them for Bangla language. In: IEEE 10th international conference on computer and information technology, pp 1–4

Umadevi KS, Chopra R, Singh N, Aruru L, Kannan RJ (2018) Text summarization of Spanish documents. In: International conference on advances in computing, communications and informatics (ICACCI), pp 1793–1797.

Vijay S, Rai V, Gupta S, Vijayvargia A, Sharma MD (2017) Extractive text summarisation in hindi. In: IEEE International conference on Asian language processing (IALP), pp 318–32

Widyassari PA, Affandy NE, Fanani AZ, Syukur A, Basuki RS (2019) Literature review of automatic text summarization: research trend, dataset and method. In: IEEE International conference on information and communications technology (ICOIACT), pp 491–496.

Yu H, Kaufman YJ, Chin M, Feingold G, Remer LA, Anderson TL, Balkanski Y, Bellouin N, Boucher O, Christopher S, DeCola P, Kahn R, Koch D, Loeb N, Reddy MS, Schulz M, Takemura T, Zhou M (2006) A review of measurement-based assessments of the aerosol direct radiative effect and forcing. Atmos Chem Phys 6(3):613–666