International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 5 May 2022, pp: 2000-2006 www.ijaem.net ISSN: 2395-5252

A Compendium of Abstractive Text Summarization

Meena Siwach[1], Jatin Rauthan[2],Sarthak Jain[3]

[1]Assistant Professor , Maharaja Surajmal Institute of Technology, Janakpuri
[2]Student,Maharaja Surajmal Institute of Technology, Janakpuri , New Delhi
[3]Student,Maharaja Surajmal Institute of Technology, Janakpuri , New Delhi
Corresponding Author : Jatin Rauthan

Submitted: 15-05-2022 Revised: 25-05-2022 Accepted: 28-05-2022

ABSTRACT:There is a huge flow of knowledge on the Internet on all the latestand old topics. A brief summary of this information is useful to many users. The text should be automatically summarized to save both user time and resources. Automatic text summarization is actually creating a short summary of a very long text and contains only meaningful and useful information about any topic. The text summary was first used in the 1950s. Since then, there has been great interest among researchers in exploring new modern ways of summarizing text so that summaries created using these techniques are consistent with summaries created by humans. Has been done. There are two ways to create a summary. 1) Abstractive Summary 2) Extractive Summary. Abstraction techniques are more complicated because they require a high level of natural dialect processing. Researchers are willing to find abstraction techniques to obtain more accurate and useful summaries. Several extraction methods have been used so far and work is underway. These methods use machine learning, deep learning, and optimization techniques. Throughout this paper, we have presented a detailed search for various written summaries of the ongoing abstraction methods. This article also describes some abstract methods. Finally, this paper concludes by explaining future areas that still have room for improvement and areas that need improvement.

KEYWORDS:Textsummarization, Deep Learning, Machine Learning, Natural dialect Processing Artificial intelligence.

I. INTRODUCTION

Text summaries automatically create a condensed form of information containing only the correct information. Ittakesonlythe important words or lines from the whole text and it should be smaller than the whole report. It was first introduced in the 1950s and since then great advances have been made in this area of research. Single and multi-document summaries are the two most common types of summaries based on the number of documents. In a single document summary, a summary is made from a single document, but in a multi-document summary, a number of documents are attached to make a summary. A single text summary can be extended to summarize multiple documents. Summarizing multiple documents, on the other hand, is more challenging than summarizing a single document. Automatic text summarization is a difficult task and requires thinking about sentence order, redundancy issues, etc. to make the resulting summary compact, on-point and relevant.Due to the proliferation of large online data streams,the demand for text summaries has increased greatlyin recent years. There are a large number of reports online and it is difficult to find universal information. There is a high chance of redundancy in the text due to the large volume of texts on a wide variety of topics. Text summarization is indispensable so that we can skip a large amount of reading text and study only the main part and save time and resources. The four main goals of this approach are information coverage, information receptivity, information redundancy, and text coherence. The summarization task can be supervised or unsupervised. In order to select important content in documents, for supervised task system requires training data to select important content. A big chunk of labeled or annotated data is necessary for reading strategies. These systems are considered at the sentence level as a problem of dividing the two categories, with abbreviated sentences selected as positive samples and non-abbreviated sentences selected as contradictory samples. Vector Support Machine (SVM) and neural networks are two popular methods of classification used for sentence separation. However,training data is not required by unsupervised systems. The summary is created only by acquiring the targeted text. As a result, for any newlydetected data unsupervised systems are appropriate without any further modification. Heuristic rules are applied by unsupervised systems to extract the most pertinent statements and then create a summary. Techniques used in unsupervised systems group is clustering. There are two types of summaries based on the output style: indicative and informative summaries. Summaries indicate what does the document concern. They provide details on the document's subject. While covering the themes, informative summaries provide all of the material in an extended way. Also,there are two types of summaries available: generic and query-specific. Query-focused summaries are also known as user-focused summaries or topic-focused.A query-related summary contains the query's content, whereas a generic summary provides a general understanding of the content present in the document.

The birth of Web 2.0 causes the creation of new types of things such as websites for social networking, forums and blogs on websites,etc... where the users can share their emotions or give their opinion on almost anythingfrom blogs, photos, videos to the service provided by a hotel, restaurant or a club etc. As a result, emotional outbursts have emerged. Text Summary(TS)and Emotional/Sentiment Analysis (SA) are the two parts of archeology that work together to form these summaries. In these abstracts, ideas are first received and categorized based on behavior (whether sentence or objective) and then polarity (positive, negative or neutral).In extract aggregation, a certain score is assigned to the lines in the report and then the high scores of the lines are selected to generate the summary. Compression rate determines the measure of summary. An abstract is created where the words or phrases are distinct from the words or phrases of the original report. It makes extensive use of natural dialect processing. This is different from the extract summary.

II. ABSTRACTIVE METHODS FOR TEXTSUMMARIZATION

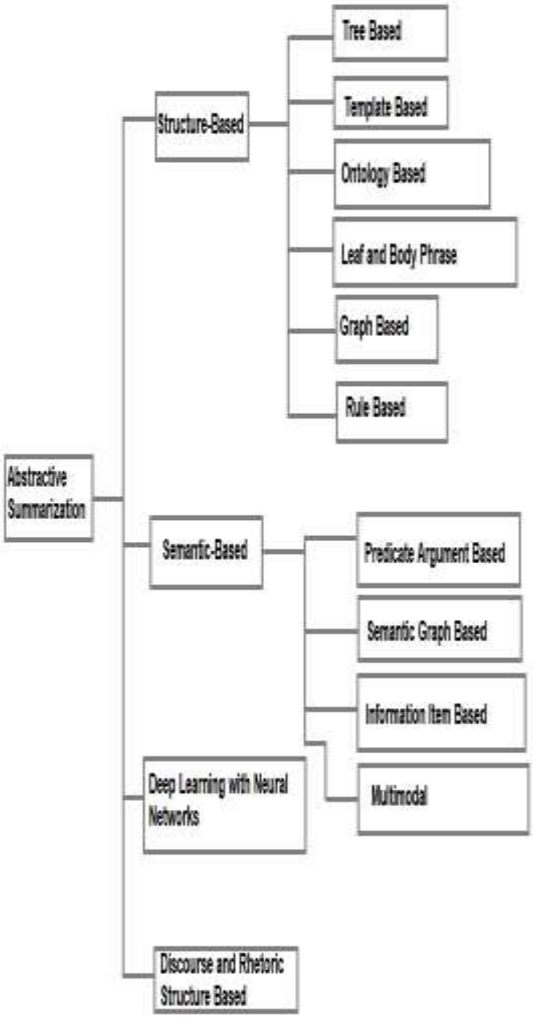


DIAGRAM ON METHODS OF ABSTRACTIVE TEXT SUMMARIZATION

Highly redundant opinion's abstractive summarization using the method of the graph

In this method, graphs are used for producing short abstractive synopsis of extremely redundant opinions. It is extremely versatile as it doesn't need any kind of domain knowledge & it employs deep NLP. In this method, firstly the graph of text is constructed, presenting the condensed text. Next, for producing the candidate's synopsis of abstraction different sub paths in the graph are traversed & marked by using 3 distinct features of the graphs. Moreover, a readable, short, well-formed & informative synopsis is produced that holds significant matter.

Abstractivetext summarization for telugu reports

Each report is pre-processed, summarized, post-processed by this method. For summarization, a variety of significant properties are employed to produce a synopsis like a keyword extraction, word clues, line selection, extraction of the line & synopsis production. Ultimately after processing, the synopsis of extraction is changed to the synopsis of abstraction by deploying synopsis refinement & rephrasing of synopsis.

Abstractivetext summarization using word graphs

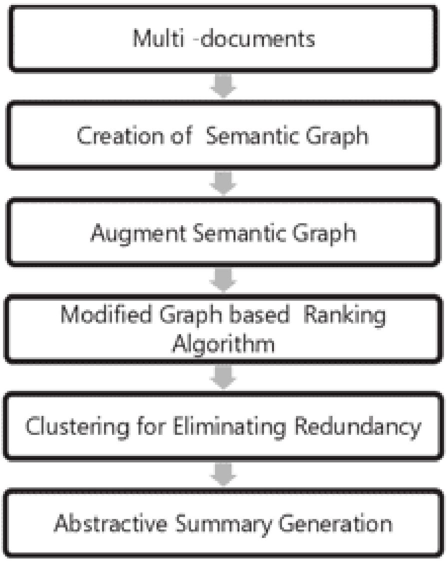
Information is Compressed & merged by this method from lines to create new lines. A text of extraction summarization method, COMPENDIUM is used for deciding which of new lines should be chosen for producing a synopsis of abstraction. Discrete methods are examined to talk about the problems related to the production of summary kind how to produce lines, order which is predominant matter can be chosen & length of lines. Findings prove that the production of the summary is a difficult task. The experiments show that by merging extraction & abstraction of data, a summary of good quality can be achieved.

Abstractive text summarization method using text-to-text production

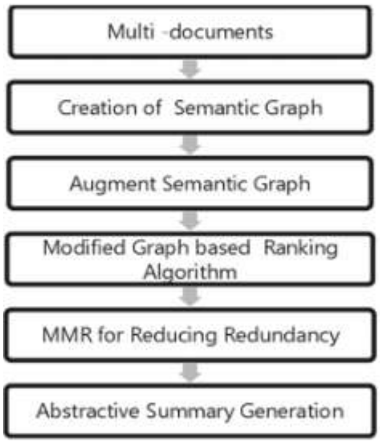
Fouroperational steps are there in this method i.e. INIT retrieval, line production, line selection & synopsis production. The matter & structure of the report is controlled by this method. A synopsis is produced by this method with a good pyramid score & linguistic quality.

Textabstractionsummarizationusingthe technique of semantic graph reduction

This method is implemented in the 3 phases: firstly production is done of the high semantic graph from original reports, next the rich semantic graph reduction thus produced to the more abstracted graph & lastly production of a summary. The original text is minimized to 50% by this approach.


Multi -documents
Creation of Semantic Graph
Augment Semantic Graph
Modified Graph based Ranking Algorithm
Clustering for Eliminating Redundancy
Abstractive Summary Generation

FIRST SEMANTIC GRAPH REDUCTION TECHNIQUE


Multi -documents
Creation of Semantic Graph
Augment Semantic Graph
Modified Graph based Ranking Algorithm
MMR for Reducing Redundancy
Abstractive Summary Generation

SECOND SEMANTIC GRAPH REDUCTION TECHNIQUE

III. COMPENDIUM, THE TEXT SUMMARIZER FOR THE EPITOME OFRESEARCH PAPERS

COMPENDIUM generates abstracts of biomedical papers. Two types of COMPENDIUM are there, COMPENDIUM$_E$ for producing extract & COMPENDIUM$_{E-A}$ which holds both methods of abstraction & extraction methods in which after selecting significant lines, data compression & stage of fusion is applied. Quality & quantity assessment of these approaches was completed and it was inferred that COMPENDIUM is appropriate for producing synopsis as both the types choose significant matter from the original report but summaries of abstractive orientation producedby COMPENDIUM$_{E-A}$ are more suitable from human perception.

Abstractive summarization of multiple reports via ilp based multiple lines together compression

Going on the most important report from the many report sets are selected in this an abstractive summarizer. Every sentence from the significant report is required to produce separate groups. Lines of reports that have the largest similarity with the group line are allocated to the group. A structure of word graph is created from lines of every group, the K-shortest paths are produced. Then lines are chosen from shortest set paths by using a new ILP problem that maximizes knowledge matter & the linguistic quality & minimizes redundancy in synopsis finally.

Selection of phases & merging based abstractive summarization of many reports

This approach uses verb or noun phrases to create new lines. Verb/Noun phrases are extracted from input reports. A prominent score is computed for every phrase by using redundancy of report matter. Phrases are chosen & merged simultaneously to achieve the solution of global optimum to create new lines whose rationality is warranted through an integer linear optimization Prototype. Other approaches are outperformed by this approach mainly on the linguistic quality assessment manually.

Summarization of abstraction using a neural attention-based prototype

Due to new inventions in the translation of neural machines, The neural attention-Prototype, which is integrated with the contextual input encoder, is created. This Prototype produces every word of the synopsis based on a given sentence. A large amount of data can be trained through it. A summarization Prototype is trained for headline production on the Gigaword dataset consisting of article pairs. Several other abstractive & extractive approaches are importantly outperformed by this prototype.

Multilingual text summarization approach through hmsm

Text Summarization is used in the multi language summary with multiple reports via the Hidden Markov Story Prototype. The summarization algorithm combines different reports into different sets. The prototype of Markov Hidden Story is directed to each set from a report group in each vernacular language using SKM recording. Viterbi goals are used to test the TDT3 collection database.English & Chinese reports can be evaluated from it.

Multilingual text summarization using dialect independent approach

Generic text is condensed through this dialect independent algorithm. Structural & statistical properties are employed through this method. It is a flexible method. The theme of the report is represented through a vector. The matter is divided& significant lines are selected from every part. English, Hindi, Gujarati & Urdu are applied through this. Summaries have a good size of symbolism in comparison to DUC synopsis For English articles. Degree symbolism is above 80% for other dialect articles other than English.

Newsgist based on statistical technique, multilingual news summarization approach

This approach to multilingual news text summarization employs the SVD (Singular Value Decomposition) technique, NewsGist. In this various documents are selected through various channels & makes important news articles from it. Summaries are produced through this method for each distinct newsgroup. Three phases are predominant in it -: interpretation, transformation & production. Line matrix is evolved in the interpretation stage for a collection of matter. The synopsis is produced by choosing only predominant lines. English, German, French are applied through it.

Text summarizer through rst & marking of lines

No machine learning is used in this text summarization method & the user can limit the length of the synopsis. Two phases are there in it :Phase 1: A primary synopsis is designed using RST. Phase 2:Each sentence of the primary synopsis isScored. Lines are selected for the final synopsis while considering that the total score of the synopsis is maximum when condensed text is under the provided limitations. RST describes the text & coherence of matter. Newspapers are evaluated through this. Farsi & Urdu can also be applied through this.

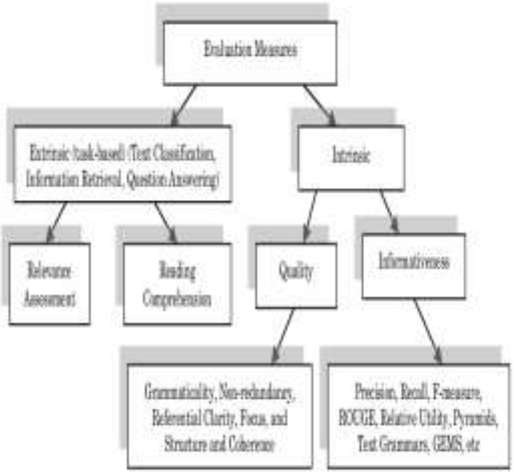
Multilingual summarizer for reports through a hybrid algorithm

This method uses properties of the Hindi text summarization approach well as the Punjabi text summarization approach. Nine features are predominant in it. For computing featured weights from the report's instruction groups we use the technique of mathematical regression. Afterward, every line marks are calculated for every feature.

Hugely pronounced lines are selected to make the shortened text.

A method for multilingual text summarization by distributed representationoftext

In this multilingual text summarization approach, important lines are selected from various reports by this method. In this method a summarization approach is followed-:(1) Lines are grouped by k-Medoid algorithms by taking aid of semantic matter available in report & (2) lines are pronounced using word's definition through this method. The technique for line comparability is moved forward to find good comparability between texts in 2 lines. This is advanced based on the necessary size of the synopsis. English, French & Arabic are executed through this approach.


Evaluation Measures
Extrinsic (task-based) (Text Classification, Information Retrieval, Question Answering)
Intrinsic
Relevance Assessment
Reading Comprehension
Quality
Informativeness
Grammaticality, Non-redundancy, Referential Clarity, Focus and Structure and Coherence
Precision, Recall, F-measure, ROUGE, Relative Utility, Pyramids, Test Grammars, GEMS, etc

Classification of summary evaluation indicators

IV. TEXT SUMMARIZATION'S FUTURE PROSPECTS

Over the last fifty years, in the field of summarization a lot of study has been done. There have been new techniques to include linguistic qualities into the summary, so it is no longer just a collection of sentences. This topic of study is constantly evolving, fulfilling new consumer needs but also encountering numerous hurdles. As a result, the emphasis in this part is on the critical concerns that have arisen in this field of research and that the research community has to address. Existing methods for summarizing text are updated periodically as text summary systems are built using modern machine learning methods. However, there is little variation in the elements required to extract relevant phrases (location, term frequency, etc.). Some new properties of words and sentences must be identified so that psychologically significant sentences can be generated from the text. There are changes in the kind of abbreviations to suit the changing needs of the user. In the beginning standard summaries of one document have been made but now due to large availability amount of data in different formats and different languages and due to the rapid development of technology, multitasking, multilingualism, multimedia abbreviations have become popular. This is noticeable in the experimental programs currently working on new kinds of summary tracks. Abbreviations with a specific goal as emotional, personal abbreviations,etc. are also produced,another crucial story is how such information can be delivered. The majority of systems nowadays deal with text input and output. New approaches could be proposed, with input in the form of meetings, films, and so on, and output in format, without text.. Other systems can be upgraded at installation is in text mode and output can be shown by tables, statistics, visual measurement scales, diagrams, etc. that allow visual effects and users have access to the required short-term content.Many new approaches to the language field have been proposed improve the level of abbreviations. Summary programs that use language methods, on the other hand, take more processor and memory space because they require greater language expertise and complex language tactics. Language resources (Lexical Chain, Context Vector Space, WordNet, etc.) and languages present extra challenges. The lack of multiple linguistic resources is as excellent as the lack of analytical instruments (analytical discussion). As a result, practical summaries based on data are required. Like a human summary, algorithms can summarize texts in any language and produce high-quality summaries. In addition to composing sentences, the content of the summary should be consistent.Therefore, the invisible or mixed path needs to be greatly improved. Important information can be chosen, collated, compressed, or additional data can be based on new summary data. By integrating output and abstractive techniques, a hybrid method can be built to produce a high-quality summary. Research continues to produce abbreviations so that machine-generated abbreviations are closely related written by people. Another significant issue is the testing system. Internal and external test methodologies are discussed in this work. Many experiments are natural which also separates information and quality assurance and performs using the latest tools and techniques. Numerous latest tools test existing details in summary and meager methods that attempt to assess the quality of the summary. There are new ways is upgraded to make the quality check process automatic by professional judges. Mostly, the accessible internal research methods focus on them is a routine vocabulary between the machine produced and the reference summary. Internal testing can be used to do research., as a result of the information it holds and its presentation, new summarizing methods are being developed. The procedure of testing is extremely reliant. The first step is to define a positive condition in order for the system to be clear about what it is important and bad. It is still unknown what he will do after leaving the post default. Similarly, summary quality testing is also highly dependent, as is the case done in person by professional judges. There are other quality testing measures such as grammar, consistency, and so on, however when the same summary is checked by two experts, the results are different. Text summarization has been around for almost 50 years, and the scientific community is particularly interested in it to continue to develop existing methods to summarize texts or develop novel summarizing techniques to produce high quality summaries. But still the functionality of the text summary is substantial and the summaries made are less complete.Therefore, this program can be made more intelligent by combining it with other programs in sequence an integrated system can do better.

V. CONCLUSION

Summary of text is an exciting field of study and offers a wide range of applications. The goal of this article is to provide academics with crucial information regarding the history of text summarization, its current state, and its potential. This article categorizes well known text abstraction approaches into different categories. Both techniques of summary testing, intrinsic and extrinsic evaluation, are briefly explored. Emphasis is mainly on the abstractive and multilingual techniques of text summarization. Finally, researchers are given some good future directions that will aid them in building summary production tactics to advance the research field.

REFERENCES

[1]. Banerjee S Mitra P, Sugiyama K (2015) Multi-document abstractive summarization using ILP based multisentence compression. In: Proceedings of the 24th international joint conference on artificial intelligence(IJCAI 2015), pp 1208–1214

[2]. Baralis E, Cagliero L, Mahoto N, Fiori A (2013) GRAPHSUM : discovering correlations among multiple terms for graph-based summarization. Inf Sci 249:96–109. doi:10.1016/j.ins.2013.06.046

[3]. Bing L, Li P, Liao Y, Lam W, Guo W, Passonneau RJ (2015) Abstractive multi-document summarization via phrase selection and. arXiv preprint arXiv:1506.01597

[4]. Chan SWK (2006) Beyond keyword and cue-phrase matching: a sentence-based abstraction technique for information extraction. Decis Support Syst 42:759–777. doi:10.1016/j.dss.2004.11.017

[5]. Fattah MA (2014) A hybrid machine learning model for multi-document summarization. 592–600. doi:10.1007/s10489-013-0490-0

[6]. Fung P, Ngai G (2006) One story, one flow: hidden Markov Story Models for multilingual multidocument summarization. ACM Trans Speech Lang 3:1–16. doi:10.1145/1149290.1151099

[7]. Ganesan K, Zhai C, Han J (2010) Opinosis : a graph-based approach to abstractive summarization of highly redundant opinions. In: Proceedings of the 23rd international conference on computational linguistics,pp 340-348

[8]. Genest PE, Lapalme G (2011) Framework for abstractive summarization using text-to-text generation. In: Proceedings of the workshop on monolingual text-to-text generation, Association for Computational Linguistics, pp 64–73

[9]. Hadi Y, Essannouni F, Thami ROH (2006) Unsupervised clustering by k-medoids for video summarization. In: ISCCSP'06 (the second international symposium on communications, control and signal processing)

[10]. Kabadjov M, Atkinson M, Steinberger J et al. (2010) NewsGist: a multilingual statistical news summarizer.Lecture notes in computer science (including including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics) 6323 LNAI, pp 591–594. doi:10.1007/978-3-642-15939-8_40

[11]. Kallimani JS, Srinivasa KG, Eswara Reddy B (2011) Information extraction by an abstractive text summarization for an Indian regional language. In: Natural language processing and knowledge engineering(NLP-KE), 2011 7th international conference on IEEE, pp 319–322

[12]. Khan A, Salim N, Jaya Kumar Y (2015) A framework for multi-document abstractive summarization based on semantic role labelling. Appl Soft Comput 30:737–747. doi:10.1016/j.asoc.2015.01.070

[13]. Lloret E, Palomar M (2011a) Analyzing the use of word graphs for abstractive text summarization. In: IMMM2011, first international conference, pp 61–66

[14]. Lloret E, Romá-Ferri MT, Palomar M (2013) COMPENDIUM: a text summarization system for generating abstracts of research papers. Data Knowl Eng 88:164–175. doi:10.1016/j.datak.2013.08.005

[15]. Moawad IF, Aref M (2012) Semantic graph reduction approach for abstractive Text Summarization. In: Proceedings of ICCES 2012, 2012 International Conference on Computer Engineering and Systems, pp132–138. doi:10.1109/ICCES.2012.6408498

[16]. Oufaida H, Philippe B, Omar Nouali (2015) Using distributed word representations and mRMR discriminant analysis for multilingual text summarization. In: Natural language processing and information systems.Springer International Publishing, pp 51–63

[17]. Rush AM, Chopra S, Weston J (2015) A neural attention model for abstractive sentence summarization. arXiv preprint arXiv:1509.00685

[18]. Meena Siwach, Suman Mann, Anomaly detection for web log data : A Survey,IEEE Conference,2022.

[19]. Meena Siwach, Suman Mann, Anomaly detection for web log data analysis using improved PCA Technique, Journal of information and optimization Science. 131-141, DOI: 10.1080/02522667.2022.2037283