International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: SJIF: 5.995
Available Online at www.journalijcar.org
Volume 7; Issue 2(C); February 2018; Page No. 9779-9783
DOI: http://dx.doi.org/10.24327/ijcar.2018.9783.1631

Research Article

LITERATURE REVIEW ON AUTOMATIC TEXT SUMMARIZATION

Reeta Rani and Sawal Tandon*

Lovely Professional University

A R T I C L E I N F O

Article History:
Received 15th November, 2017
Received in revised form 21st December, 2017
Accepted 23rd January, 2018
Published online 28th February, 2018

Key words:

Automatic Text Summarization (ATS), Natural Language Processing (NLP), Data Mining (DM), Abstractive Summary, Extractive Summary.

A B S T R A C T

Automatic text summarization (ATS) is nowadays a popular research area among researchers. Automatic text summarization is the approach of generating the subset of the main text. This subset of the main text represents the complete text and the main idea of the text. Automatic Text summarization is also known as Test summarization. ATS is the important field of Natural Language Processing (NLP) and Data Mining (DM). This includes the abstractive and extractive summaries of the text. This review paper provides the overview of various past researches and study in the field of Automatic Text Summarization.

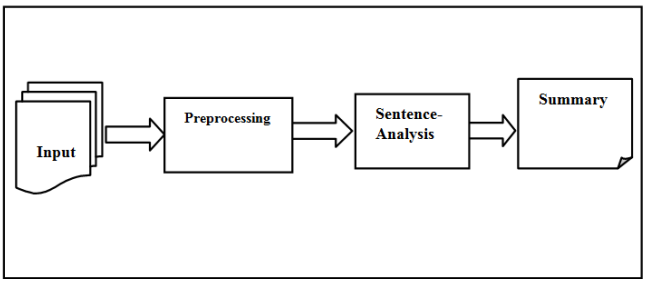
INTRODUCTION

With the advancement of technology, the internet is accessible through various devices like smartphones, smartwatches and within the reach of common people. That leads to the accessibility of lot of information through world wide web (WWW). More information on the internet is in text form so sometimes it becomes so difficult to select only required information from large texts. Due to the large volume of information, manual summarization of information is very challenging and also time-consuming task [1]. Thus, we need an automatic text summarization system. Radev et al [2] defined Summary as "a text that is produced from one or more texts, that conveys the important information in the original text(s), and that is no longer than half of the original text(s) and usually significantly less than that".

Summaries make the task of understanding the meaning of text easier. Text summarization helps user to manage vast amount of information by condensing document and include more relevant facts into them [3]. Text summarization process contains three steps: analysis, transformation, and synthesis [4]. The Analysis step analyzes the The Fig.1 shows the general steps of text summarization or ATS. The input of the system can be single or multiple documents. It depends on the user requirement. The next step is Preprocessing in this step stop words removed and tokenization performed.

Corresponding author:* **Sawal Tandon
Lovely Professional University

Sentence Analysis step includes sentence scoring and sentence ranking to rank the sentence. From this, the final summary is generated which is the final and the last step of the system.



[19] **Fig.1** General steps of Text Summarization

Automatic text summarization technology is also able to summarize multiple documents and then presents the summary of all multiple documents into one summary [6]. Such summaries are used for summarizing the multiple news from different sources and reviewing products. Search engines can also take the advantage of automatic text summarization to provide the summarize information of large web pages to users. The rest of the paper is organized as follow: Section 2 contains the different challenges in text summarization. Section 3 contains the classification of text summarization techniques. Section 4 contains literature review and in Section 5 we discussed afinal conclusion.

Challenges

As the text summarization provides the reduced subtext of the original text there are always a lot of challenges there to measure if we get the required text summary from the large

document or not. If, the summary contains important sentences and words or not. Some of the challenge of text summarization that is also called as automatic text summarization are discussed as follow:

Extract hidden semantic relationship

In the text, there are many sentences that are related to each other. The text includes many sentences. Some sentences in the document related to each based on have semantic relationships between concepts in the text [7]. Many sentences are related to each other that depict some particular details of the text document. So, capturing such sentences in the summary is always challenging task.

Relevance detection

While generating a summary of the text it is also important to find the relevant sentences in our text documents so that they can be selected to make up the final summaries [8]. This is always a challenging task to find out the relevant sentences in text so that they can be included in our final summary. As it highly affects the quality of our summary. "The Code Quality Principle" can be used to detect important sentences in our text document [8] [9]. Different criteria's like *sentence position within the text* [10], *word and phrases frequencies* [11] [12] and *title overlap* [13] are some of the examples to ensure the relevance of the sentence.

Summary Evaluation

Theevaluation of generated summaries is also the challenging task the summaries can be evaluated either manually that is so hard or either by using some automatic text summary evaluation methods which is again a difficult task. The difficulty is arising due to the lack and impossibility of building the gold-standard against which we can compare the final results [14]. It is difficult to determine good summary because there is always thepossibility that the system generated summary is different from human-generated summaries. There are two approaches suggested for evaluation of summaries Intrinsic evaluation and extrinsic evaluation [15]. An automatic summarization ROUGE tool is also used for automatic summary evaluation [16].

Classification of Text Summarization Techniques

There are different techniques of text summarization. All the broadly classified techniques can be grouped into following three groups [17].

Extraction Based Summarization Techniques

In extraction based summarization the sentences and keywords are selected that can be included in the summary. The sentences and words are selected from the main text. In this, the main text is divided into the sentences or words and select and reject them according to significance score. The selected sentences and words are included in the final summary.

Abstraction Based Summarization Techniques

In abstraction based summarization the new sentences are generated that represent certain text information from the main text. We accomplish this through the natural language generation. This generates more human like information in which sentences represents the main idea of the text in summary rather the including exact sentences or words from the main text in the final summary like Extraction based summarization techniques.

Aided Summarization Techniques

In aided summarization techniques uses the machine learning techniques and neural networks. The machine learning techniques like Naive Bayes classifier and support vector machine. These are combined with the existing data mining. This helps in classification of sentences and words if they can be included in final summary after trained on the data sets.

LITERATURE REVIEW

Chin-Yew Lin [16] In this paper author introduced Recall-Oriented Understudy for Gisting Evaluation ROUGE. That is an automatic evaluation package for text summarization. The paper also introduced four different measures of ROUGE: - ROUGE-N, ROUGE-L, ROUGE-W and ROUGE-S. It measures the quality of summary by comparing the generated summary with other ideal summaries that are created by humans. These methods are efficient for automatic evaluation of single document summary as well as multi-document summaries.

Akshil Kumar et al. [17] In this paper author has analyzed and compared the performance of three different algorithms. Firstly, the different text summarization techniques explained. Extraction based techniques are used to extract important keywords to be included in the summary. For comparison three comparison three keyword extraction algorithms namely TextRank, LexRank, Latent Semantic Analysis (LSA) were used. Three algorithms are explained and implemented in python language. The ROUGE 1 is used to evaluate the effectiveness of the extracted keywords. The results of the algorithms compared with the handwritten summaries and evaluate the performance. In the end, the TextRank Algorithm gives abetter result than other two algorithms.

Pankaj Gupta et al. [18] In this paper author has reviewed different techniques of Sentiment analysis and different techniques of text summarization. Sentiment analysis isa machine learning approachin which machine learns and analyze the sentiments, emotions present in the text. The machine learning methods like Naive Bayes Classifier and Support Machine Vectors (SVM) are used.these methods are used to determine the emotions and sentiments in the text data like reviews about movies or products. In Text summarization, uses the natural language processing (NPL) and linguistic features of sentencesare used for checking the importance of the words and sentences that can be included in the final summary. In this paper, asurvey has been done of previous research work related to text summarization and Sentiment analysis, so that new research area can be explored by considering the merits and demerits of the current techniques and strategies.

Harsha Dave et al. [19] In this paper author has proposed a system to generate the abstractive summary from the extractive summary using WordNet ontology. The multiple documents had been used like text, pdf, word files etc. The author has discussed various text summarization techniques then author discussed step by step the multiple document text summarization approaches. The experiment result is compared with the existing online extractive tools as well as with human-generated summaries and shows the proposed system gives good results. At last the author proposed for the future that the

summarization accuracy can be increasedby comparing this abstractive system with some other abstractive system.

Yihong Gong et al. [20] In this research paper the author proposes two methods that create the generic text summaries by ranking and extracting sentences from the main text documents. The first method uses information retrieval (IR) methods that rankthe sentence relevance and provides the relevance scores to sentencesand thesecond method uses the latent semantic analysis (LSA) technique that based on latent semantic indexing (LSI) in order to identify the semantic importance of the sentences, for summary creations. The author uses the Singular Value Decomposition (SVD) to generate the text summary. Further, this paperauthor explainsthe SVD based methods step by step. The effect of different Weighted Schemes is also checked on the performance of the summaries. The purposed methods provide generic abstractive summaries. Finally, the results are compared with the human-generated summaries. It generates better human like abstractive summaries. For future author proposed to investigate various machine learning techniques so that quality of generic text summarization can be improved.

Opinosis. It generates abstractive summaries. Opinosis works on redundant data like human reviews on movies or products and provides abstractive summaries. Firstly, it creates the direct Opinosis-Graph of the text. Where nodes represent the word units of the text. Three unique graph properties: Redundancy capture, Collapsible structures and Gapped subsequence capture is used to explore and explore different sub-paths that help in the creation of abstractive summaries of the text. The valid path is selected and marked with high redundancy score, collapsed path and summary generation. Then all paths ranked in descending order according to scores. The duplicate paths are removed using Jaccard measure the results are compared with human summaries. Results show Opinosis summaries has better agreement with human summaries. For future work author proposed to use a similar idea to overlay parse trees.

Dharmendra Hinhu et al. [24] In this paper the author uses the extractive text summarization. The author gives the Wikipedia Articles as input to the system and identifies text scoring.

Table I

Author	Year	Techniques/Methods	Outcome
Chin-Yew Lin *et al*.	2004	Graph based approach	Abstractive Summary generation of redundant data
Akshil Kumar *et al*.	2017	Graph based approach, Semantic based approach	Performance of Three different algorithms compared TextRank, LexRank and LSA. TextRank outperforms other two.
Pankaj Gupta *et al*.	2016	Sentiment Analysis, Text Summarization Techniques	A Survey is performed on current research in sentiment analysis and Text summarization.
Harsha Dave *et al*	2015	Ontology based	Generated abstractive summary from extractive summary
Yihong Gong *et al*.	2001	Semantic based	New LSA method provides generic text summary.
Rada Mihalcea *et al*.	2004	Graph based	New TextRank method generates extractive text summary.
Güneş Erkan *et al*.	2004	Graph based	New LexRank method generates extractive text summary
Kavita Ganesan *et al*.	2010	Graph based	New framework Opinosis generates abstractive summary of redundant data
Dharmendra Hinhu *et al*.	2015	Extractive Text summarization approach	Extractive text summarization approach is used to summarize Wikipedia Articles.
N. Moratanch*et al*.	2016	Structure based, Semantic based Approaches	A Survey on various techniques of Abstractive text summarization.
Tacho Jo	2017	K- Nearest Neighbor	Modified KNN provides Text summarization

Rada Mihalcea et al. [21] In this paper the author introduced the TextRank a graph-based ranking model for the processing of the text. it is an unsupervised method for keyword and sentence extraction. TextRank uses voting based weighting mechanism and provides the score to the sentence then finally determine the importance of the sentence. The nodes in the graph represent the sentences. The significance of the sentence based on incoming and outgoing edges from nodes. The weight of each is determined based on similarity score between the sentences. TextRank derived from the Google's Page Rank algorithm. TextRank provides extractive summaries of the text. Text Rank Provides the best results.

*Güneş Erkan et al. [22]*In this paper the author introduces graph-based method LexRank. In this, the sentence score is calculated based on Eigenvector Centrality. It is cosine transform weighting method. In this, the original text is split into sentences and a graph is built where sentences act as the nodes. The complete method is explained in the paper. The results show that LexRank outperforms the existing centroid-based methods. This method is also performed well in case of noisy data. This method generates an extractive summary of the text.

*Kavita Ganesan et al. [23]*In this research paper the author proposed graph-based text summarization framework

Firstly, the sentences are Tokenized through pattern matching using regular expressions. Then we get data in form of set of words then stop words are removed from the set of words. The words are then stemmed. Then traditional methods are used for scoring of the sentences. Scoring helps in classifying the sentences if they included in summary or not. It is found that scoring sentences based on citation give better results.

*N. Moratanch et al. [25]*In this paper the author presents an exhaustive survey on abstraction based text summarization techniques. The paper presents a survey on two broad abstractive summary approaches: Structured based abstractive summarization and Semantic-based abstractive summarization. The author presents the review of various researches on both approaches of abstractive summarization. The author also covered the various methodologies and challenges, in abstractive s summarization.

*N. Moratanch et al. [26]*In this paper the author presents the comprehensive review of extraction based text summarization techniques. In this paper the author provides survey on extractive summarization approach by categorized them in: Supervised learning approach and Unsupervised learning approach. Then different methodologies, the advantages are presented in the paper. The author also includes various

evaluation methods, challenges and future research direction in the paper.

Tacho Jo [27] In this paper the author proposed a particular version of KNN (K Nearest Neighbor) where the words are assumed as features of numerical vectors represents text. The similarity between feature vectors is computed by considering the similarity among attributes as well as among values. Text summarization viewed as the task of classification. The text is partitioned into paragraphs or sentences. Each paragraph or sentence is classified into 'summary or 'nonsummary' by the classifier. The sentences which are classified into 'summary' are extracted as results from summarizing the text and other text rejected. Improved results are obtained with the proposed version of KNN in text classification and clustering. The modified version of KNN leads to a more compact representation of data item and better performance.

CONCLUSION

As the availability of data in the form of text increasing day by day. It becomes so difficult to read the whole textual data in order to find the required information which is both difficult as well as a time-consuming task for a human being. So, at that time ATS performs an important role by providing a summary of a whole text document by extracting only the useful information and sentences. There are different approaches of text summarization. The real-world applications of text summarization can be: documents summarization, news and articles summarization, review systems, recommendation systems, social media monitoring, survey responses systems. The paper provides a literature review of various research works in the field of automatic text summarization. This research area can be explored more by looking in existing systems and working on different and new techniques of NPL and Machine Learning.

References

1. SaiyedSaziyabegum, Priti S. Sajja, "Literature Review on Extractive Text Summarization Approaches" *International Journal of Computer Applications* (0975-8887) Volume 156- No 12, December 2016.
2. Kang Wu, Ping Shi, Da Pan, "An Approach to automatic summarization for Chinese text based on the combination of spectral clustering and LexRank." IEEE Access 2016.
3. Ibrahim F. Moawad, Mostafa Aref, "Semantic graph reduction approach for abstractive Text Summarization." *Seventh International Conference on Computer Engineering & Systems* (ICCES), 2012.
4. K.Sparck Jones, "Automatic Summarising: The State of the Art" Information Processing & Management, vol. 43, pp. 1449-1481, Nov 2007.
5. Bhavana Lanjewar," Automatic text summarization withcontext-based keyword extraction ", *International Journal of Advance Research in Computer Science and Management Studies*, Vol. 3, Issue 5, May 2015.
6. GlorianYapinus, Alva Erwin, MaulahikmahGaliniu, WahyuMuliady, "Automatic Multi-Document Summarization for Indonesian Documents Using Hybrid Abstractive- Extractive Summarization Technique". 6th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, 2014.
7. S.ABabara, Pallavi D. Patilb, "Improving Performance of Text Summarization". International Conference on Information and Communication Technologies ICICT, 2014.
8. Elena Lloret and Manuel Palomer, "Challenging issues of automatic summarization: relevance detection and quality-based evaluation." Informatica 34, no. 1, 2010.
9. T. Givón, T. "Isomorphism in the Grammatical Code: Cognitive and Biological Consideration." In R. Simone (ed.). 47-79, 1994.
10. H.P.Edumundson, "New Methods in automaticExtracting." In: Inderjeet Mani and Mark Maybury, editors, Advances in Automatic Text Summarization, MIT Press pp. 23-42, 1969.
11. H.P. Luhn, "The Automatic Creation of literature abstracts." In: Inderjeet Mani and Mark Maybury, editors, Advances in Automatic Text Summarization, MIT Press pp. 15-22, 1958.
12. E. Lloret, O. Ferrández, R. Muñoz, M. Palomar, "A Text summarization Approach Under the Influence of Text Entailment." In: Proceeding of the 5th International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2008) 12-16 June, Barcelona, Spain. 22-31, 2008.
13. D.R.Radev, S. Blair-Goldensohn, Z. Zhang, "Experiment in Single and Multi-Document Summarization using MEAD." In: First Document Understanding Conference, New Orleans, LA. 1-7, 2001.
14. M. Fuentes Fort, "A Flexible Multitask Summarizer for Document from Different Media, Domain and Language." Ph.D. thesis (2008) Adviser-Horacio Rodríguez.
15. I. Mani, "Summarization Evaluation: An Overview." In: Proceeding of the North American Chapter of the Association for Computational Linguistics (NAACL). Workshop on Automatic Summarization, 2001.
16. Chin-Yew Lin, "Rouge: A Package for Automatic Evaluation of Summaries." Barcelona Spain, Workshop o Text Summarization Branches Out, Post- Conference Workshop of ACL 2004.
17. Akshi Kumar, Aditi Sharma, Sidhant Sharma, Shashwat Kashyap, "Performance Analysis of Keyword Extraction Algorithms Assessing Extractive Text Summarization." International Conference on Computer, Communication, and Electronics (Comptelix), 2017.
18. Pankaj Gupta, RituTiwari and Nirmal Robert, "Sentiment Analysis and Text Summarization of Online Reviews: A Survey." International Conference on Communication and Signal Processing, 2016.
19. Harsha Dave, Shree Jaswal, "Multiple Text Document Summarization System using Hybrid Summarization Technique." 1st International Conference on Next Generation Computing Technology (NGCT), 2015.
20. Yihong Gong, Xin Liu, "Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis." Proceeding of the 24th annual international ACM SIGIR Conference on research and development in information retrieval, ACM 2001.

21. Radha Mihalcea, Paul Tarau, "TextRank: Bring Order into Texts." Association for Computational Linguistics, 2004.
22. GüneşErkan, Dragomir R. Radev, "LexRank: Graph-based Lexical Centrality as Salience in Text Summarization." *Journal of Artificial Intelligence Research* 22 457-479, 2004.
23. Kavita Ganesan, ChengXiangZhai, Jiawei Han, "Opinosis: A Graph-Based Approach to Abstractive Summarization of Highly Redundant Opinions." Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pages 340-348, 2010.
24. DharmendraHingu, Deep Shah, Sandeep S.Udmale, "Automatic Text Summarization of Wikipedia Articles." International Conference on Communication, Information & Computing Technology (ICCICT), 2015.
25. N. Moratanch, Dr. S. Chitrakala, "A Surveyon Abstractive Text Summarization." International Conference on Circuit, Powe and Computing Technologies (ICCPCT), 2016.
26. N.Moratanch, S. Chitrakala, "A Surveyon Extractive Text Summarization." IEEE International Conference on Computer, Communication and Signal Processing (ICCCSP), 2017.
27. Taeho Jo, "K Nearest Neighbor for Text Summarization using Feature Similarity." International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), 2017.

How to cite this article:

Reeta Rani and Sawal Tandon (2018) 'Literature Review on Automatic Text Summarization', *International Journal of Current Advanced Research,* 07(2), pp. 9779-9783. DOI: http://dx.doi.org/10.24327/ijcar.2018.9783.1631

****