RESEARCH REVIEW JOURNALS
Volume-03
Issue-05
May-2018

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal]

Review on Abstractive Text Summarization Techniques for Biomedical Domain

[*1]Krutika Patel & [2]Urmi Desai

[*1]*Computer Engineering Department, Sarvajanik College of Engineering and Technology, Athwalines, Surat, Gujarat (India)*
[2]*Computer Engineering Department, Sarvajanik College of Engineering and Technology, Athwalines, Surat, Gujarat (India)*

ARTICLE DETAILS

Article History
Published Online: 04 May 2018

Keywords
Abstractive Summary, Text Summarization, rich semantic graph

***Corresponding Author**
Email: kruti2443[at]gmail.com

ABSTRACT

In this internet era amount of biomedical literature and data are increased exponentially. In order to keep up to date with knowledge of this field and other related area information also interpret the outcome of experiments in light of all available literature, researchers turn more and more to the use of automated literature mining. Biomedical or Biological domain is all about studying life and tremendous amount of biomedical textual information has produced and collected all over the world on daily basis. The task of analyzing huge amount of biomedical data and association of biological data is much difficult. To efficiently analyze the biomedical domain data text summarization approach is used. Automatic text summarization provides solution by generating summary automatically. Text summarization techniques classified into extractive and abstractive text summarization types. Existing techniques of extractive text summarization extract important sentences from original document and generate summary without any modification of actual data. This technique may not present conflicting information properly. Abstractive text summarization can solve this problem by representing the extracted sentences into another understandable semantic form. This paper discusses abstractive text summarization techniques and highlights the parametric evaluation of these techniques.

1. Introduction

Text Summarization is a process of reducing data from the vast amount of literatures. For biomedical field tremendous amount of information are there for clinical and researchers from a variety of sources, for example, scientific literature databases, Electronic Health Records (EHR) systems, web archives, patient's reports and interactive media records. The scientific literatures give wellsprings of data to researchers like MEDLINE, PubMed, IEEE and ACM digital library. Clinical trials and scientific publications supply a new researches or technology frequently for more advancement in biomedical field. It helps the clinicians and researcher analysts to look for important information and save their time to seek information. Some reasons have been identified for producing summaries from full-text documents even when they provide abstracts. The reason incorporates there are variants of an ideal summary in addition to the abstract, 1) some content of the full-text may be missed in the abstract, 2) customized summaries are useful in question answering systems, 3) automatic summaries allow abstract services to scale the number of documents they can evaluate, and 4) assessing the quality of sentence selection methods can be helpful in development of multi document summarization system.

Automatic text summarization gives a decent intends to fast obtaining of data through compression and refinement. While existing strategies for automatic text summarization achieve elegant performance on short sequences, however, they are facing the challenges of low efficiency and accuracy when dealing with long text. An automatic text summarization is an effective technique, which utilizes computers to process and compress texts in order to produce concise and refined content. In the time of enormous information and rapid of information overload, automatic text summarization has become an important and timely tool for user to quickly understand the large volume of information.

The automatic summarization is the core subtle part of natural language processing [1][3].Automatic text summarization used in many areas, for example, news articles outlines, email summary, short message news on portables, information summary for businessman, online search engines and biomedical and so forth [8][9][11].

In extractive text summarization extracted sentences could become longer than the average [2][3]. Due to this some of the portion which are not important for summary that also gets included. Moreover the conflicting information may not be presented properly [2]. Abstractive text summarization can solve this problem by representing the extracted sentences into another more understandable semantic form [2]. In this paper we are studying different techniques of abstractive text summarization.

This paper aims to make survey of existing abstractive text summarization techniques along with parametric evaluation of these techniques.

This paper covers the details: various different text summarization techniques described in section 2. The parametric evaluation of abstractive text summarization techniques is presented in section 3. Finally, section 4 concludes with a discussion of future research directions in this area.

2. Related Work

This section gives a detailed description of various abstractive text summarization techniques. Depending on the

input and other parameters summarization categorized into two group's extractive and abstractive summarization.

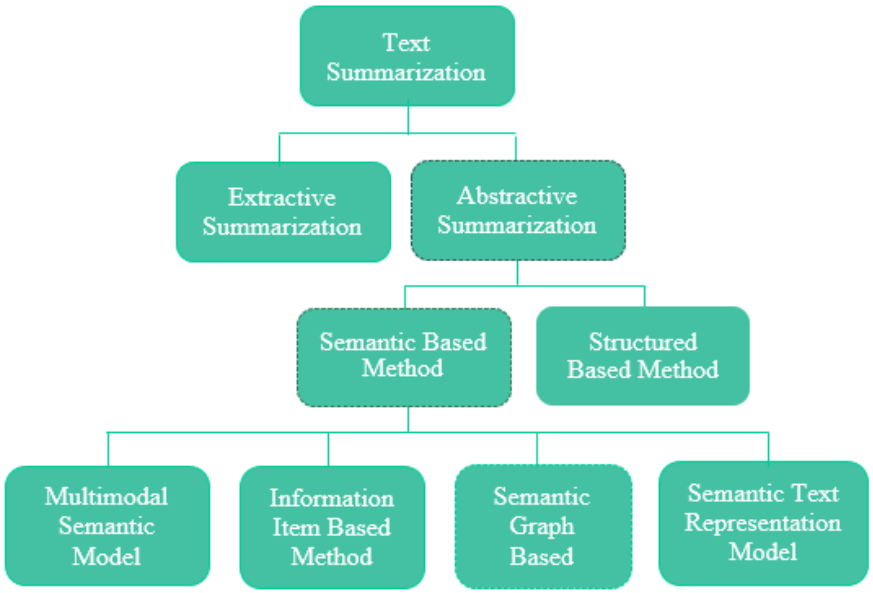


Fig. 1.0: Text Summarization Basic Process

Abstractive summarization classified into two categories: structured base and semantic base. In Abstractive Semantic based method, semantic representation of document(s) used to feed into natural language generation (NLG) system. This method focus on identifying noun phrases and verb phrases by processing linguistic data. Different techniques utilized this approach are discussed here [10][14][16]:

Multimodal semantic model catches the ideas and relationship among source information that important ideas evaluated based on some measures and finally the selected concepts expressed as sentences to form summary. In information item based method the contents of summary generated from abstract representation of source documents. The abstract representation called information item, in which the smallest element of coherent information in a text. In Semantic graph based method summary of document forms by creation a rich semantic graph (RSH) of the original document, reduced the generated semantic graph, and then generating the final abstractive summary. Semantic text representation method analyzed input text using semantics of words rather than syntax structure of text.

3. Various Techniques for Text Summarization

Most of work done in text summarization has focused in this section, we discuss different approaches and some works on abstractive text summarization.

A. Semantic Graph Reduction Approach

This approach [1] outlines an input document by creating semantic graph called as rich semantic graph (RSG). The semantic graph further reduced and generates final abstractive summary from reduced semantic graph. System takes input as a solitary document in English language and output generated as reduced summary report.

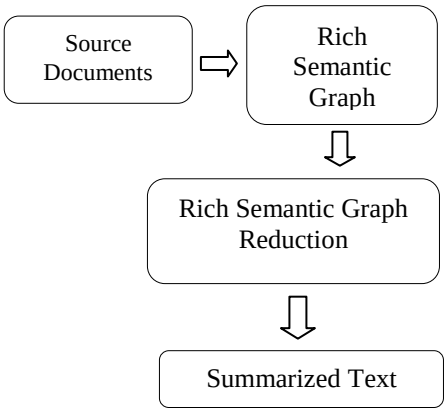


Fig. 2.0: Architecture of Semantic Graph Reduction Approach

This approach comprises of three task. The first task is RSG creation. The main aim of the RSG creation to represent the input document semantically. In that verb and nouns of input document represented as graph nodes and edges represented as semantic relations between them. In this way it builds the graph for each sentence and afterward it interconnects rich semantic sub-graphs. At the end the sub-graphs, all the sub-graph merged together to represent the whole document semantically. The second phase called RSG reduction. In this phase a set of rules applied on RSG to reduce it by merging, deleting the graph nodes. Third phase generates abstractive summary from reduced RSG. This approach succeeds to reduce the source document up to half of the original document. Limitation of this approaches that no multiple documents taken as input to generate abstractive summary.

B. Word Graph based Approach

This approach uses word graph to represent source document. This approach includes two phases [2]. First phase sentence reduction and second sentence combination. The sentence reduction phase based on discourse rules to remove

redundant clauses at the beginning of a sentence, and syntactic constraints to complete the end of the reduced sentence. Word graph used for sentence combinations and to represent word relations between texts [12]. New sentences are generated from several sentences which are generated by using word graph. In word graph nodes used to store the information about words and their part of speech tagger and edges used to represent adjacency relations between word pairs. This approach generate syntactically correct sentence but does not care about word meaning.

C. Sentiment Infusion Approach

This approach work on a graph based technique that generates summaries of redundant opinions and uses sentiment analysis to combine the statements. Also uses word graph for compressing and merging information and then summaries are generated from resultant sentences. The graph captures the redundancy in the document using words that occur more than once in the texts that mapped to the same node. Moreover, the graph creation does not require any domain knowledge. At the time of graph generation this approach will ensure the correctness of sentences.

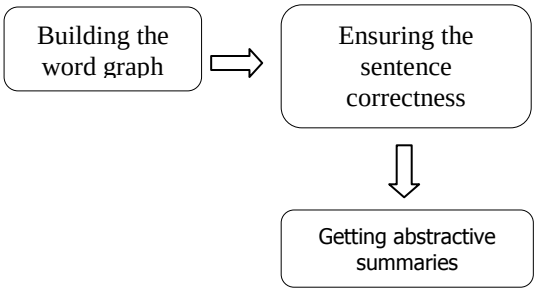
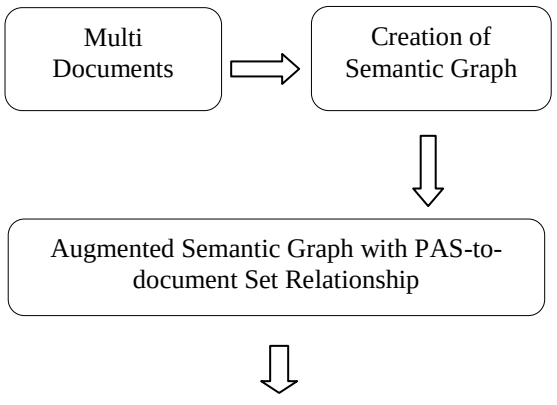


Fig. 3.0 Architecture of Sentiment Infusion Approach [3]

For getting abstractive summary, score given to all the paths as well as the sentences have been fused. After that ranked the sentences in descending order of their scores and remove duplicate sentences from summary using jacquard index for similarity measure. Then the remaining top most sentences chosen for the summary.

D. Genetic Semantic Graph Based Approach

This approach [4] work on a genetic semantic graph based approach for multi document text summarization. This approach constructs a semantic graph from document text in such a way that the graph nodes represent the Predicate



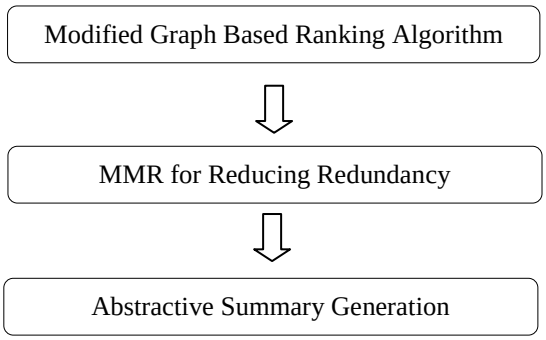


Fig. 4.0 Proposed Genetic Semantic Graph Based Approach [3]

Argument Structure (PASs) and the edges of the graph represent a semantic similarity weight which can be determined from PAS-to-PAS semantic similarity, and PAS-to-document set relationship. For constructing PASs they use semantic role labelling.

In order to reduce redundancy, utilize maximal marginal relevance (MMR) to re-ranks the PASs and use language generation to generate summary sentences from the top ranked PASs [13]. This approach automatically merges similar information across the documents to reduce the overlapping information in summary.

E. Clustered Genetic Semantic Graph based Approach

This [5] work on clustered genetic semantic graph based approach for multi document abstractive text summarization. This approach similar to genetic semantic graph based approach but that used clustering algorithm to eliminate redundancy. Algorithm eliminate redundancy in such a way that representative PAS with the highest similarity score from each cluster chosen and fed to language generation to generate summary sentences. For making cluster used Hierarchical Agglomerative Clustering (HAC) algorithm [15]. HAC algorithm accepts the semantic similarity matrix as input. Algorithm merges two clusters which most similar and update the semantic similarity matrix to represent the pair wise similarity between the nearest cluster and the original cluster.

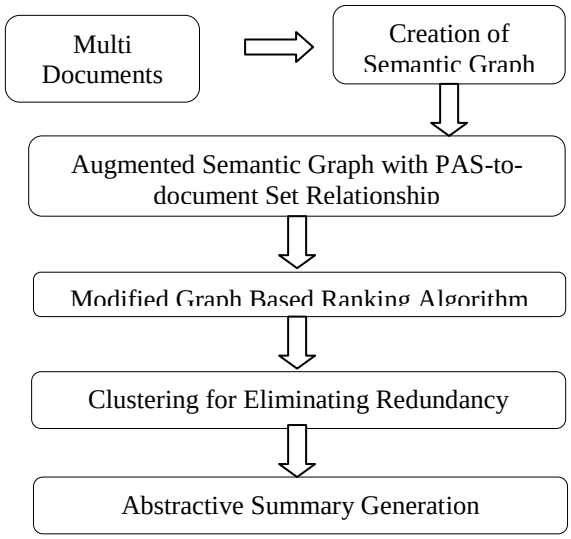


Fig. 5.0 Proposed Clustered Genetic Semantic Graph Based Approach [5]

Process repeats until the compression rate of summary reached 20%. Once the clusters obtained, top scored PASs obtained using simple natural language generation and a simple heuristic rules form to generate summary sentences from PASs.

4. Comparison of abstractive text summarization techniques

This section illustrates comparison of previously discussed abstractive text summarization techniques use full for biomedical domain [12] [14]. Table 1 shows a comparative study of abstractive text summarization techniques based on parameters as follows. Type of text summarization parameter indicates that abstractive summary to be generated from single source document or multi documents. Source document representation parameter constituted that the original text to be represented in which form. Content selection parameter represent that which techniques or algorithm used for extracting important information. Summary generation parameter describes that final abstractive summary generated in which form. Semantic summarization parameter and syntactically correct representation parameter indicates that generated summary is semantically and syntactically correct or not. This all techniques are based on mono-lingual language based techniques. There are other languages also available like multi-lingual and cross-lingual. In mono-lingual language based technique, input and output language is same. However in multi-lingual language input document would be in more than one language and output will be in the user desired language and in cross-lingual language, input and output language is different from each other.

Table 1. Parametric Evaluation of Abstractive Text Summarization Techniques

Technique	Type of Text Summarization	Original Text Representation	Content Selection	Summary Generation	Semantically Correct Summarization	Syntactically Correct Representation	Technique used for Eliminate Redundancy
Title Semantic Graph Reduction Approach [1]	Single document	Rich semantic graph	Heuristic rules	Reduced semantic graph	Yes	No	-
Word Graph based Approach [2]	Single document	Word graph	Relation among words, Clauses	Word graph	No	Yes	-
Sentiment Infusion Approach [3]	Single document	Word graph	Sentiment analysis	Path Scoring, Sentence Fusion	Yes	Yes	-
Genetic Semantic Graph based Approach [4]	Multi document	Semantic graph	Semantic Role Labeling and Semantic Similarity Score	SimpleNLG and Simple Heuristic rule	Yes	Yes	Maximal Marginal Relevance (MMR) algorithm
Clustered Genetic Semantic Graph based Approach [5]	Multi document	Semantic graph	Semantic Role Labeling and Semantic Similarity score	SimpleNLG and Simple Heuristic rule	Yes	Yes	Clustering algorithm

5. Conclusion

In this paper, we study different abstractive text summarization techniques based on natural language processing, data mining and semantic similarity approaches. These all techniques used to generate summary automatically from source document. This, All techniques are mono lingual language based. Semantic graph based reduction approach produces concise, coherent and less redundant sentences. Sentiment infusion approach generates summary which semantically and syntactically correct and in reduced form. Among all text summarization techniques, clustered genetic semantic graph based approach eliminate the overlapping semantic redundancy significantly.

Future work may include developing a more efficient technique with multi- lingual or cross-lingual structure based. One can also try to generate more concise and less redundant summary by designing new approach or by merging available techniques which provides accurate summary results for biomedical domain.

References

1. IF. Moawad, M. Aref, “Semantic Graph Reduction Approach for Abstractive Text Summarization, Computer Engineering & Systems (ICCES)”, *IEEE*, 2012, pp. 132-138.
2. H. T. Le, T. M. Le, “An approach to abstractive text summarization, Soft Computing and Pattern Recognition (SoCPaR)”, IEEE, 2013.

3. R. Bhargava, Y. Sharma, G. Sharma, “ATSSI: Abstractive Text Summarization using Sentiment Infusion, Procedia Computer Science”, Elsevier, 2016.
4. A. Khan, N. Salim and Y. J. Kumar, “Genetic Semantic Graph Approach for Multidocument Abstractive Summarization, Digital Information Processing and Communications(ICDIPC)”, IEEE, 2015.
5. A. Khan, N. Salim and H Farman, “Clustered Genetic Semantic Graph Approach for Multi-document Abstractive Summarization”, Intelligent Systems Engineering (ICISE)”, *IEEE* , 2016.
6. A. R. Pal, D. Saha, “An approach to automatic text summarization using WordNet, Advance Computing Conference (IACC)”, IEEE, 2014.
7. K. S. Thakkar, R. V. Dharaska, “Graph-Based Algorithms for Text Summarization”, Emerging Trends in Engineering and Technology (ICETET)”, IEEE, 2010.
8. J. Zhan, H. T. Loh, Y. Liu, “Gather customer concerns from online product reviews – A text summarization approach”, Expert Systems with Applications, Elsevier, 2009.
9. T. Workman, M. Fiszman and J. Hurdle, "Text summarization as a decision support aid", *BMC Medical Informatics and Decision Making,* vol. 12, no. 1, 2012.
10. H. Thanh, T. Manh, “An approach to Abstractive Text Summarization”, IEEE, 2013.
11. R. Mishra, J. Bion and M. Fiszman, “Text summarization in the Biomedical Domain: A Systematic Review of Recent Research”, *J Biomed Inform,* 2014.
12. T. Workman, M. Fiszman and J. Hurdle, "Text summarization as a decision support aid", *BMC Medical Informatics and Decision Making*, vol. 12, no. 1, 2012.
13. H. Menendez and L. Plaza and D. Camacho, "Combining graph connectivity and genetic clustering to improve biomedical summarization", *IEEE Congress on Evolutionary Computation,* Beijing, China, 2016.
14. H. Thanh, T. Manh, “An approach to Abstractive Text Summarization”, IEEE, 2013.
15. I. Yoo, X. Hu and I. Song, "A coherent graph-based semantic clustering and summarization approach for biomedical literature and a new summarization evaluation method", *BMC Bioinformatics*, vol. 8, no. 9, p. S4, 2007.
16. H. Reeve, H. Han and D. Brooks, “The use of domain-specific concepts in biomedical text summarization”, *ELSEVIER*, 17 July 2006.

