

Application of Different Algorithms to Optimize Abstractive Summarization



Bhumika, V. Srikanth

Abstract: Summarization is used to extract most relevant content from a huge content. It can be extractive and abstractive. But different people may have different scope of relevance in different area. Content can be in the form of opinions, news, judgments and ideas etc. Extractive Summarization extracts most focusing content without any change in the original content. Abstractive summarization is a knowledgebase extraction with some modification in original content. In this paper first we discuss various algorithms for abstractive summarization then on the basis of merits of various algorithms we discuss, how various algorithm may help to optimize Abstractive Summarization.

Challenges in Abstractive summarization:

- Repetition or duplicity of the content
- Arrangement of words
- Different people have different meaning for the same content
- Number of times a content repeated
- Different people have different interest
- Nosiness in content
- Information diversity

Methods for Abstractive summarization:

1. integer linear programming-based summarization framework [3]
2. Improved Semantic Graph Approach [4]
3. novel speech act-guided summarization approach [5]
4. Neural Abstractive Summarization with Diverse Decoding(NASDD) [6]
5. Supervised and unsupervised approach [7]
6. Neural network based representation [8]
7. Integer linear optimization [9]
8. Maximum L_{∞} -norm and minimum entropy regularization [10]
9. Opizer-E and Opizer-A [11]
10. LSTM-CNN based ATSDL model [12]
11. novel concept-level approach [13]

Keywords: Abstractive Summarization, Diverse Decoding [6], Opinion Summarization [11].

I. INTRODUCTION

In today’s fast-moving life no one have time to explore the things in detail, everyone needs everything in summarized form. Same content is available on different web pages. The ‘Redundant Big Data’ need summarization tool for its maximal utilization. This paper will discuss various summarization techniques used for text, document, bilingual or multilingual document, opinion etc. It will also discuss some advanced graphical and voice recognition techniques used for summarization. Some parts of this paper will also cover the advanced and automated technologies used for summarization like Neural Network, Voice over search, fuzzy set and Deep learning. These techniques make the summarization process more effective and efficient.

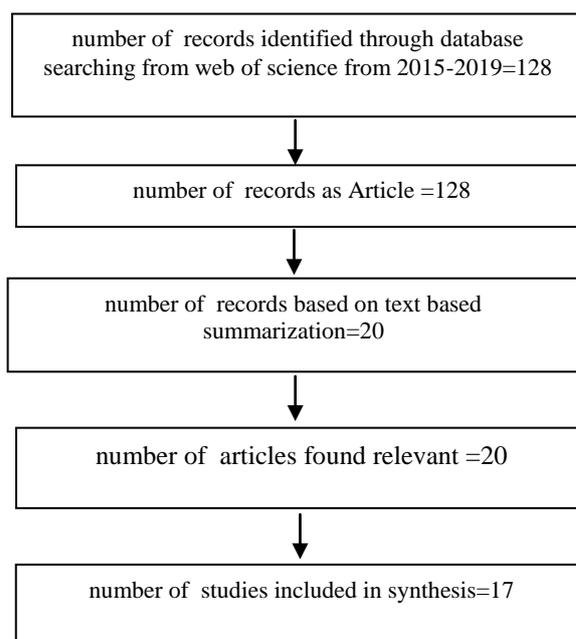
1. Extractive Summarization: Extracting significant sentences, organizing and making another sentence utilizing unique words from which it is separated.
2. Abstractive Summarization: Restructuring important substance utilizing new words utilizing semantic examination.

II. METHODOLOGY USED

A. Methodology

A comprehensive computerized database search of all type documents, from (2015-2019) was carried out, using database “Web of science”.

B. Block Diagram 1



Manuscript published on January 30, 2020.

* Correspondence Author

Bhumika*, Assistant Professor, Department of Computer Applications, Dr. Y.S. Parmar Govt. PG College Nahan, District Sirmaur, Himachal Pradesh, India.

Dr.V.Srikanth, Associate Professor, Department of AIT IBM CSE Chandigarh University, Panjab, india

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)



III. ALGORITHMS FOR ABSTRACTIVE SUMMARIZATION AVAILABLE

A. integer linear programming-based summarization framework [3]:

Different people may have different Lexical choices for same Events. E.g. People may have different opinion related to same events. This type of problem called lexical diversity. To Overcome Lexical Diversity ILP based Frame Work is used. ILP extracts most focusing sentences from the source documents without any duplicity.

Co occurrence matrix checks concept based resemblances and optimize it using the proximal gradient method.

B. Improved Semantic Graph Approach [4]:

Most graph method is based on measure of similarity of content and ignores semantically equivalence of sentences. In semantic graph method node represent predicate argument structure, edges represent similarity weight. This method mostly focuses on frequency of sentences.

Algorithm: Ranking algorithm on the basis of graph

1) Initialization

Set position scores of all vertices to 1.

2) Computation of rank score

2.1) Find vertices of the chart.

2.2) For every vertex of the chart decide adjoining vertex.

2.3) For each adjoining vertex discover number of active edges from that vertex, and afterward summarize the loads related with those edges.

2.4) Compute the rank score of vertex considered in Step 2.2,

2.5) Update the rank score of the vertex.

2.6) Repeat Steps 2.2 to 2.5 until intermingling is accomplished.

3) Sort the rank scores of diagram vertices: Sort the rank scores backward request.

Formal setting investigation (FCA) strategy can use for building idea pecking order from content consequently. At that point significance might be estimated.

C. novel speech act-guided summarization approach [5]

Content present on twitter can be of arbitrary length, can be related to different fields and contains lots of noise. Words that are present with higher frequency are higher in ranking. On the basis of that we can generate template-based summaries.

Vocal acts can be used to extract important keywords and phrases.

It is proposed that keywords classifiers can be improved using Naïve Bayes and Decision Tree.

D. Neural Abstractive Summarization with Diverse Decoding (NASDD) [6]

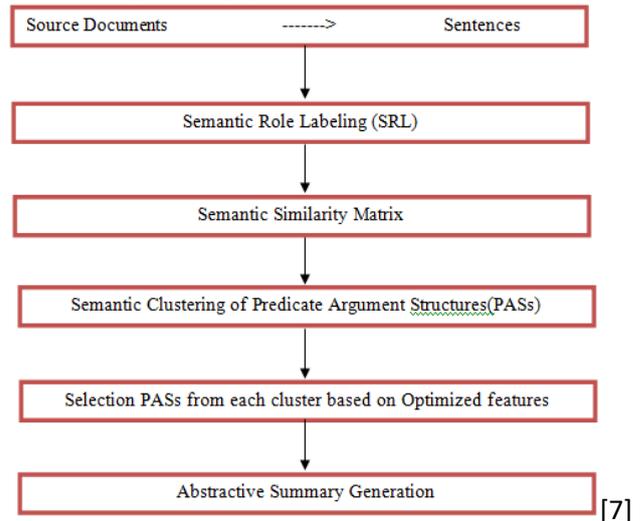
Diversity-promoting beam search approach increases the possibility of generating more relevant summary in the decoding process. Attention mechanism combined with the key features can cover salient information, which aids in finding the optimal summary sequence.

Diversity-Promoting Beam search (DPBS) may produce outputs with different perspectives. Diversity at discourse level can generate multiple summaries with a single input.

E. GA with semantic role labeling [7]

Genetic Algorithm summarizes multiple documents and having following steps:

Diagram 2 [7]



F. Supervised and unsupervised approach [15]

This algorithm covers:

(i) Sentiment examination incorporates commentary, order, and discovering connection between prime highlights and audit outlines.

(ii) Supervised approach with the diagram based technique can be utilized for sentiment outline.

(iii) Both composed content and Vocal guides can create yield for outline.

This technique is utilized to abridge unconstrained discussions. In solo technique exchange structure with chart based strategy correspondence is evaluated; in managed approach remarkable quality score, conclusion score, theme importance and sentence length are taken as area of assessment.

G. Neural network based representation [16]

Natural language understanding helps to learn and generate semantic structures more effectively.

H. Integer linear optimization [9]

Cross-Language Summarization is also possible by Translation Model Enhanced Predicate Argument Structure and Fusing.

In this algorithm a translation system is used to convert source language document into targeted language document. New sentences are generated by introducing some external bilingual words as PAS elements using ILP algorithm to improve semantic structure.

I. Maximum L_{∞} -norm and minimum entropy regularization [10]

Consideration on catchphrases can be improved utilizing L_{∞} -standard regularization and least entropy regularization.

There are two sorts in meager consideration model, first is the L_{∞} -standard regularization amplifies the consideration on most centering words at set loads of every single other word to 0.0.

Least entropy regularization hones the utilization of different words. This model is successful for accomplishing significant enhancements for interpretation execution and outline quality.

Later on, they will utilize neural systems based outline for consideration based substance.

J. Opizer-E and Opizer [11]-

The conclusions present on web about any item, spot, and friends and so on requires outline.

The conclusion rundown proposes two new strategies for outline named as Opizer-E (an extractive technique) and Opizer (an abstractive strategy).

The sentiment outline strategy utilizes Natural Language Generation (NGL) layout bases framework to create abstractive rundowns of suppositions.

Result shows that extractive strategies are superior to the abstractive techniques for assessment outline.

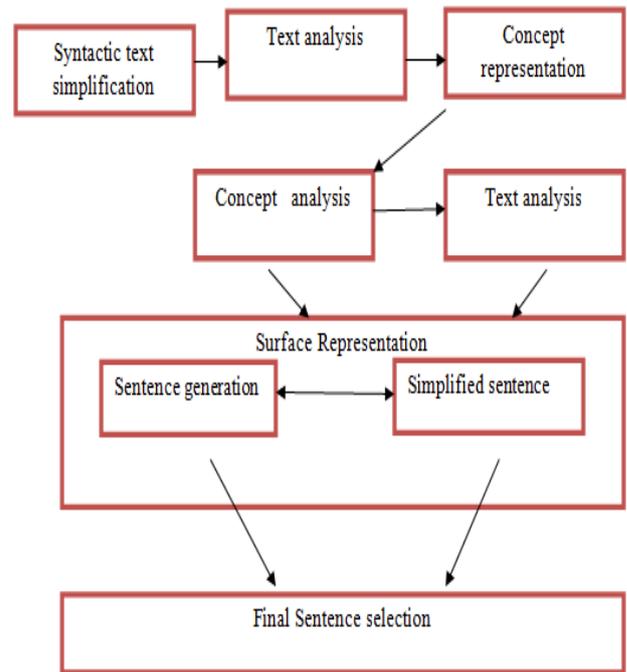
K. LSTM-CNN based ATSDL model [12]

ATSDL includes two levels, in first level we extricates phrases from source sentences and the second creates content synopses utilizing deep learning.

L. novel concept-level approach [13]

The fundamental thought behind this methodology as for other existing algorithm approaches is the incorporation of an improvement and a recovery stages.

Diagram 3[13]



[13]

IV. RESULT AND DISCUSSION

Table- I: Ideas used by various Authors

Title	Year	Author	Method	Description	Merits
A novel ILP framework for summarizing content with high lexical variety [3].	2018	Luo, Wencan Liu, Fei Liu, Zitao Litman, Diane [3].	integer linear programming-based summarization framework [3].	Sentence-concept co occurrence matrix is created and optimized using the proximal gradient method [3].	1. ILP based synopsis produces an outline by choosing a lot of sentences without redundancy, at that point sentence-idea co event network is made and check whether sentences have a place with comparable idea and streamline it utilizing proximal inclination technique. 2. ILP extricate lesser substance regarding number of sentences.
Abstractive Text Summarization based on Improved Semantic Graph Approach [4].	2018	Khan, Atif Salim, Naomie Farman, Haleem Khan, Murad Jan, Bilal Ahmad, Awais Ahmed, Imran Paul, Anand [4].	Improved Semantic Graph Approach [4].	In semantic graph method node represent predicate argument structure, edges represent similarity weight. This method mostly focuses on frequency of sentences. For better results ranking algorithm is used [4].	1. Semantic graph methods can be used to reduce human intervention. 2. Ranking method can used to create domains to avoid human intervention.
Automatic Twitter Topic Summarization With Speech Acts [5]	2018	Zhang, Renxian Li, Wenjie Gao, Dehong Ouyang, You [5]	novel speech act-guided summarization approach [5]	Generate template-based summaries, 'templates designed for speech acts, using a round-robin algorithm [5].	Extraction of keywords and phrases can be improved by vocal method to compose abstractive summaries and to control noise.

Application of Different Algorithms to Optimize Abstractive Summarization

Title	Year	Author	Method	Description	Merits
Diverse Decoding for Abstractive Document Summarization [6]	2019	Han, Xu-Wang Zheng, Hai-Tao Chen, Jin-Yuan Zhao, Cong-Zhi [6]	Neural Abstractive Summarization with Diverse Decoding(NASDD) [6]	First, they introduce a diversity-promoting beam search approach; Second, we creatively utilize the attention mechanism combined with the key information [6].	Diverse Decoding System can be used to found optimal summary sequence.
A framework for multi-document abstractive summarization based on semantic role labelling [7]	2015	Khan, Atif Salim, Naomie Kumar, Yogan Jaya [7]	GA with semantic role labelling [7]	Semantic role labelling (SRL) technique Approach is used to create template [7].	All Abstractive summarization is dependent on human being for the construction of features and rules. But this limitation can be overcome by using Genetic algorithm in which Semantic Role Labeling is done.
A novel concept-level approach for ultra-concise opinion summarization [13]	2015	Lloret, Elena Boldrini, Ester Vodolazova, Tatiana Martinez-Barco, Patricio Munoz, Rafael Palomar, Manuel [13]	novel concept-level approach [13]	Simplification and a regeneration stages [13].	Extractive summarization is better than Abstractive for opinion summarization.
Opinion summarization on spontaneous conversations [15]	2015	Lopez Condori, Roque Enrique Salgueiro Pardo, Thiago Alexandre [15]	supervised and unsupervised approach [15].	Classification and an analysis of relationship between subjective sentences and opinion summaries with the graph-based method [15].	Opinions present on internet about any product, place, company uses both extractive and abstractive summarization using NLG both qualitative and quantitative measures are applied.
Abstractive Cross-Language Summarization via Translation Model Enhanced Predicate Argument Structure Fusing [9].	2016	Zhang, Jiajun Zhou, Yu Zong, Chengqing [9]	Integer linear optimization [9].	Predicate Argument Structure and Fusing. [9]In cross language summarization we can generate summary in the targeted language by merging multiple bilingual PAS structures using integer linear optimization that attempts to maximize the salience score and the translation quality simultaneously.	Integer Linear Optimization can be used for summarization of multi-lingual Document.
Attention with Sparsely Regularization for Neural Machine Translation and Summarization [10].	2019	Zhang, Jiajun Zhao, Yang Li, Haoran Zong, Chengqing [10]	Maximum L_∞ -norm and minimum entropy regularization [10]	Applied in any attention-based sequence to sequence learning framework [10].	Sparse model is useful where attention-based extraction is required.
Abstractive text summarization using LSTM-CNN based deep learning [12].	2018	Song, Shengli Huang, Haitao Ruan, Tongxiao [12].	LSTM-CNN based ATSDL model [12].	First extracts phrases from source sentences and then generates text summaries using deep learning [12].	This method takes the advantage of using deep learning.

V. CONCLUSION

All above points discussed in results and discussions may help to refine the results of abstractive summarization. Inclusion of merits of all the algorithms may effectively increase the significance of the summary. This paper suggests how merits of different algorithm can be used to optimize the results of abstractive summarization in following manner:

- 1) Semantic graph methods can be used to optimize summarization.
- 2) Relevance, linguistic quality and usefulness are used summarization.
- 3) ILP based synopsis produces an outline by choosing a lot of sentences without redundancy, at that point sentence-idea co event network is made and check whether sentences have a place with comparable idea and streamline it utilizing proximal inclination technique.
- 4) ILP extricate lesser substance regarding number of sentences.
- 5) Extractive summarization is better than Abstractive for opinion summarization.
- 6) Ranking method can used to create domains to avoid human intervention.
- 7) Extraction of keywords and phrases can be improved by vocal method to compose abstractive summaries and to control noise.
- 8) Diverse Decoding System can be used to found optimal summary sequence.
- 9) All Abstractive summarization is dependent on human being for the construction of features and rules. But this limitation can be overcome by using Genetic algorithm in which Semantic Role Labeling is done.
- 10) Sparse model is useful where attention-based extraction is required.

- 11) Opinions present on internet about any product, place, company uses both extractive and abstractive summarization using NLG both qualitative and quantitative measures are applied.

REFERENCES

1. R. Belkebir and A. Guessoum, "Concept generalization and fusion for abstractive sentence generation," *Expert Syst. Appl.*, vol. 53, pp. 43–56, Jul. 2016.
2. Z. Cao, C. Chen, W. Li, S. Li, F. Wei, and M. Zhou, "TGSUM: Build Tweet Guided Multi-Document Summarization Dataset," in *SOCIAL MEDIA CONTENT ANALYSIS: NATURAL LANGUAGE PROCESSING AND BEYOND*, vol. 3, Wong, KF and Gao, W and Xu, R and Li, W, Ed. 2018, pp. 401–417.
3. W. Luo, F. Liu, Z. Liu, and D. Litman, "A novel ILP framework for summarizing content with high lexical variety," *Nat. Lang. Eng.*, vol. 24, no. 6, pp. 887–920, Nov. 2018.
4. A. Khan et al., "Abstractive Text Summarization based on Improved Semantic Graph Approach," *Int. J. Parallel Program.*, vol. 46, no. 5, pp. 992–1016, Oct. 2018.
5. R. Zhang, W. Li, D. Gao, and Y. Ouyang, "Automatic Twitter Topic Summarization with Speech Acts," in *SOCIAL MEDIA CONTENT ANALYSIS: NATURAL LANGUAGE PROCESSING AND BEYOND*, vol. 3, Wong, KF and Gao, W and Xu, R and Li, W, Ed. 2018, pp. 347–373.
6. X.-W. Han, H.-T. Zheng, J.-Y. Chen, and C.-Z. Zhao, "Diverse Decoding for Abstractive Document Summarization," *Appl. Sci.*, vol. 9, no. 3, Feb. 2019.
7. A. Khan, N. Salim, and Y. Jaya Kumar, "A framework for multi-document abstractive summarization based on semantic role labelling," *Appl. Soft Comput.*, vol. 30, pp. 737–747, May 2015.
8. H. Li, "Deep linear autoencoder and patch clustering-based unified one-dimensional coding of image and video," *J. Electron. Imaging*, vol. 26, no. 5, Sep. 2017.
9. J. Zhang, Y. Zhou, and C. Zong, "Abstractive Cross-Language Summarization via Translation Model Enhanced Predicate Argument Structure Fusing," *IEEE-ACM Trans. AUDIO SPEECH Lang. Process.*, vol. 24, no. 10, pp. 1842–1853, Oct. 2016.
10. R. E. Lopez Condori and T. A. Salgueiro Pardo, "Opinion summarization methods: Comparing and extending extractive and abstractive approaches," *Expert Syst. Appl.*, vol. 78, pp. 124–134, Jul. 2017.
11. S. Song, H. Huang, and T. Ruan, "Abstractive text summarization using LSTM-CNN based deep learning," *Multimed. Tools Appl.*, vol. 78, no. 1, pp. 857–875, Jan. 2019.
12. E. Lloret, E. Boldrini, T. Vodolazova, P. Martinez-Barco, R. Munoz, and M. Palomar, "A novel concept-level approach for ultra-concise opinion summarization," *Expert Syst. Appl.*, vol. 42, no. 20, pp. 7148–7156, Nov. 2015.
13. M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 1–66, Jan. 2017.
14. D. Wang and Y. Liu, "Opinion summarization on spontaneous conversations," *Comput. SPEECH Lang.*, vol. 34, no. 1, SI, pp. 61–82, Nov. 2015.
15. J. Yao, X. Wan, and J. Xiao, "Recent advances in document summarization," *Knowl. Inf. Syst.*, vol. 53, no. 2, pp. 297–336, Nov. 2017.
16. B. Pandey, P. K. Bhanodia, A. Khamparia, and D. K. Pandey, "A comprehensive survey of edge prediction in social networks: Techniques, parameters and challenges," *Expert Syst. Appl.*, vol. 124, pp. 164–181, Jun. 2019.

AUTHORS PROFILE



Bhumika, working as Assistant Professor (Computer Applications) in Dr. Y.S. Parmar Govt. PG College Nahan, District Sirmour(173001) with 5 years of teaching experience. Had Completed MCA from Panjab University Chandigarh. **Email:** profbhumika@gmail.com



Dr. V. Srikanth, is currently working as Associate Professor in Department of AIT IBM CSE Chandigarh University, having 18 years of teaching experience. Had Completed MCA, M.phil and PhD. **Email:** srikanth.e8714@cumail.in