Abstractive Summarization with NLP Makes Easier to Memorize

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 algorithm for abstractive summarization then on the basis of merits of various algorithm discussed in conclusions, we propose a new algorithm for abstractive summarization with the help of NLP to generate shortcuts from the content that make easier to memorize the content.

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
- Noitness 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].
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 world many students are competing to get Govt. Jobs. For which lot of content has to be remembered. Many psychological researches has proved that we learn by repetition. As we know cannot repeat the whole content every time. So we need to create shortcuts to memorize the content. Lot of efforts are needed to create shortcuts. In today’s fast-moving life no one have time to explore the things in detail, everyone needs everything in summarized form.

II. PROPOSED ALGORITHM

Step 1) Input Content to be summarized.
Step 2) split the content in words.
Step 3) Extracts the first alphabet from each word.
Step 4) Combine alphabets to make a meaningful word by selecting phrases from Hinglish (Hindi + English) database.
Steps 5) Arrange words using semantic analysis.
Step 6) Generate output as sentence.
Abstractive Summarization with NLP Makes Easier to Memorize

III. 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 2

IV. 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 1 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 3 [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
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.

- 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 run downs of suppositions. Result shows that extractive strategies are superior to the abstractive techniques for assessment outline.

J. 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.

K. 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 4 [13]
## V. RESULT AND DISCUSSION

<table>
<thead>
<tr>
<th>Title</th>
<th>Year</th>
<th>Author</th>
<th>Method</th>
<th>Description</th>
<th>Merits</th>
</tr>
</thead>
<tbody>
<tr>
<td>A novel ILP framework for summarizing content with high lexical variety [3].</td>
<td>2018</td>
<td>Luo, Wencan Liu, Fei Liu, Zitao Litman, Diane [3]</td>
<td>integer linear programming-based summarization framework [3].</td>
<td>Sentence-concept co-occurrence matrix is created and optimized using the proximal gradient method [3].</td>
<td>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 extract lesser substance regarding number of sentences.</td>
</tr>
<tr>
<td>Abstractive Text Summarization based on Improved Semantic Graph Approach [4].</td>
<td>2018</td>
<td>Khan, Atif Salim, Naomie Farman, Haleem Khan, Murad Jan, Bilal Ahmad, Awais Ahmed, Imran Paul, Anand [4].</td>
<td>Improved Semantic Graph Approach [4].</td>
<td>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].</td>
<td>1. Semantic graph methods can be used to reduce human intervention. 2. Ranking method can used to create domains to avoid human intervention.</td>
</tr>
<tr>
<td>Diverse Decoding for Abstractive Document Summarization [6]</td>
<td>2019</td>
<td>Han, Xu-Wang Zheng, Hai-Tao Chen, Jin-Yuan Zhao, Cong-Zhi [6]</td>
<td>Neural Abstractive Summarization with Diverse Decoding(NASDD) [6]</td>
<td>First, they introduce a diversity-promoting beam search approach [6]; Second, we creatively utilize the attention mechanism combined with the key information [6].</td>
<td>Diverse Decoding System can be used to found optimal summary sequence.</td>
</tr>
<tr>
<td>A framework for multi-document abstractive summarization based on semantic role labelling [7]</td>
<td>2015</td>
<td>Khan, Atif Salim, Naomie Kumar, Yogan Jaya [7]</td>
<td>GA with semantic role labelling [7]</td>
<td>Semantic role labelling (SRL) technique Approach is used to create template [7].</td>
<td>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.</td>
</tr>
</tbody>
</table>
VI. 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 extract 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

Abstractive Summarization with NLP Makes Easier to Memorize


AUTHORS PROFILE

**Bhumika**, working as Assistant Professor (Computer Applications) in Dr. Y.S. Parmar Govt. PG College Nahan, District Sirmaur(173001) with 5 years of teaching experience. Completed: MCA from Panjab University Chandigarh. Email: profbhumika@gmail.com, Contact No: 7018005305, 9418492860.

**Dr.V.Srikanth**, is currently working as Associate Professor in Department of AIT IBM CSE Chandigarh University, with 18 years of teaching experience. He has completed MCA, M.phil and PhD. Email: srikanth.e8714@cumail.in, Contact No: 79853 98168, 96951 59581.