International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 11 | Nov -2020 ISSN: 2582-3930

A Gentle Introduction of Text Summarization: The review

Devika Bhise[1], Sneha Vishawabramha[2]

[1]*CSE,SGGSIE&T,Nanded*
[2]*CSE,SGGSIE&T,Nanded*

--- ***---

Abstract - In the digital world everyday lots of data get added. Human summaries are varying human to human. It is needed to manage everyday created data and summaries it to save reading time, to find headlines, important short notes. The paper giving review of text summarization methods. Understanding natural language is one of the big-problem. Here, short we are giving a short introduction about text summarization. Multiple methods are available for calculating the text summarization. We are introducing some of them.

***Key Words*:**ROUGE, Abstractive, Extractive

1.INTRODUCTION

Text summarization can be defined as, it is subset of the input provided. It is a problem of getting fluent, short and accurate summary. We can also say that it is automatically generated summary of natural language. The text summarization is broadly divided into two types: Extractive summarization and Abstractive summarization.

Extractive Summarization: It is just like copy and pasting the text. In this type of summarization, the text gets shortened but we do not get the meaningful summary.
Abstractive Summarization: In this type of summarization we shortened the text by creating new meaningful sentences. It is hard problem because in this problem the new sentences are get created and synthesized.

1.1 Benefits of text summarization:
1.Text summarization reduces reading time. 2.We get important facts easily.
3.Important information gets high-lighted and we don't miss the information. 4.It makes selection process easier. 5.Notes are always helping and reducing reading efforts.

1.2 Challenges of text summarization or some issues regarding text summarization:
1.Recognizing and copying with redundancy 2. Identifying important differences among documents
3.Ensuring summary coherence even when material stems from different source documents. 4.Evaluation issues are Form matrices and content matrices.

1.3 Single-document vs. Multi-document summarization

While performing single document summarization, with some repetition of factual data system rely on unified piece of text.
Multi document system maximizes the important facts included in the summary while minimizing repetition of data.
The probability of redundancy increases with multi-document summarization systems.

2.APPROACH: There are many approaches can beused to implement text summarization, such as prediction based, frequency-based approach. Below is basic text summarization overflow.

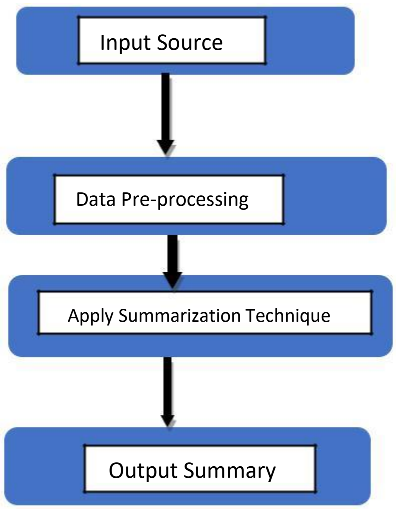

Fig -1: Basic text summarization approach

2. Survey Table

The table below, will give overall information about the available methods about the text summarization. It will help the new author for selecting algorithms, give knowledge about present methods and new challenges in this domain. It will knowledge about the present work and the work done by other authors.

References	Extractive vs Abstractive	Deep Or Algorithm	Single Vs multi	Data Set	Evalu - ation	Publi- cation &Year	Reduction Or compression (if mention)	Summary Unit
Mahnaz Koupaee William Yang Wang [1]	Abstractive	Neural- Intra attentional model	Multi	Modified CNN/daily Mail, NYT	Rouge	arXiv, 2017	Upto 75 Words	Sentences
John M.Conroy Dianne P.O'Leary[2]	Abstractive	Hidden Morkov model	-	The TREC dataset	F1 Score	Research- Gate, 2001	-	Sentences
Sandeep Sripada Venu Gopal Kasturi Gautam Kumar Parai [3]	Extractive	Stack-decoder algo novel graph problem-based formulation	Multi	DUC2004 Dataset	Rouge[6]	Stanford	Vertices 60 Edges 1400	Document
Dr.M. Suman Tharun Maddu A. Shalini Bhavna [4]	Abstractive	Using NLTK	Sing- le	-	-	Compusoft 2015	-	Sentences
Rada Mihalcea and Paul Tarau [5]	Extractive	Iterative Graph-based algorithms	Both	DUC2002 (TeMario summarization test collection	META	Aclweb, 2005	25-30%	Document
Chandra Khatri Gvanit Singh Nish Parikh [6]	Both	Recurrent Neural Networks and Document Context Vector	Sing- le	Human Extracted Snippets eBay, Wikipedia and DUC summarization task	Rouge accuracy, Precision, Recall, F-Score	arXiv, 2018	According to optimization algorithm	Sentences
Jianpeng Cheng Mirella Lapata [7]	Extractive	Neural net Work	Sing- le	Daily mail news articles, DUC2002	Rouge	arXiv, 2016	About 30%	Sentences
Joel Larocca Neto alexander D.Santos Celso A.Kaestner Alex A Freita [8]	Extractive	Summarization and auto class algorithm	Sing- le	NIHCL corpus	Compared with other System Such as SUMMAC model, CGI/CMU Systems	Research Gate,2000	Up to 10 sentences	Sentences
Ramesh Nallapati, Feifei Zhai∗ ,Bowen Zhou [9]	Extractive	Neural network based	Sing- le	CNN/Daily Mail corpus	Rouge	arXiv, 2016	Rouge at 75 word	Sentences
Abigail See Peter J.Liu Google Christopher D.Manning [10]	Both	Sequence -to - sequence attentional model	Sing- le	CNN/Daily Mail	Rouge	arXiv, 2017	35% of the time	Sentences

Ahmet Aker Trevor Cohn Robert Gaizausk [11]	Extractive	A* search algorithm	Multi	Image Description Corpus	Rouge And manual Readabili Ty	Aclweb, 2010	-	Sentences
Anshuman pattanaik,Santwana Sagnika,madhabananda das and Bhabani sankar Prasad Mishra [12]	Extractive	Bat algorithm	Single	-	Rouge	Springer, 2019	35% compression	Sentences
Chi Zhang,Lei Zhang chong-jun Wang,Jun-Yuan Xie [13]	Extractive	Sentence selection with Semantic Representation	Multi	DUC2006, DUC2007	Rouge	IEEE,2014	Restricted upto 250 words	Sentences
Ramesh Nallapati Bowen Zhou Cicero dos Santos Caglar G Ulcehre Bing Xiang [14]	Abstractive	Attentional encoder decoder model	Single	Gigaword Corpus DUC corpus, CNN/daily Mail Corpus	Rouge	arXiv, 2016	30 words for every summary	Sentences

Table -2.1: Survey Table

3. CONCLUSIONS

Fast increasing in growth of data can be handled by using a powerful summarizer. Due to rapid growth of data on social media, digital platform, text summarization still a big challenge. Getting perfect summary and evaluating it as human generated is still hard problem for machine. We are collecting is some methods and work done by other user for ease of use to new users.

REFERENCES

1. Mahnaz Koupaee University of California, Santa Barbara koupaee@cs.ucsb.edu William Yang Wang "WikiHow: A Large- Scale Text Summarization Dataset" arXiv,18 Oct 2018
2. John M. Conroy, Dianne P. O'Leary "Text Summarization via Hidden Markov Models", ResearchGate ,2001
3. Sandeep Sripada, Venu Gopal Kasturi, Gautam Kumar Parai "Multi-document extraction-based Summarization",stanford
4. Dr. M. Suman, Tharun Maddu, A. Shalini, 4K. Bhavana "A New Approach for Text Summarizer" COMPUSOFT, An international journal of advanced computer technology, 4 (4), April-2015 (Volume-IV, Issue-IV)
5. Rada Mihalcea and Paul Tarau "A Language Independent Algorithm for Single and Multiple Document Summarization", aclweb,2015
6. Chandra Khatri Gyanit Singh Nish,Parikh "Abstractive and Extractive Text Summarization using Document Context Vector and Recurrent Neural Networks",arXiv,2018
7. Jianpeng Cheng Mirella Lapata "Neural Summarization by Extracting Sentences and Words", arXiv,2016
8. Joel Larocca Neto Alexandre D. Santos Celso A.A. Kaestner Alex A.Freita "Document Clustering and Text Summarization" researchgate,2000
9. Ramesh Nallapati, Feifei Zhai∗, Bowen Zhou "SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents", arXiv,2016
10. Abigail See, Peter J. Liu, Christopher D. Manning "Get To The Point: Summarization with Pointer-Generator Networks", arXiv, 2017
11. Ahmet Aker, Trevor Cohn, Robert Gaizauskas "Multi-document summarization using A* search and discriminative training", aclweb, 2010
12. Anshuman Pattanaik, Santwana Sagnika, Madhabananda Das and Bhabani Sankar Prasad Mishra, "Extractive Summary: An Optimization Approach Using Bat Algorithm", Springer,2019
13. Chi Zhang, Lei Zhang, Chong-Jun Wang, Jun-Yuan Xie,"Text Summarization Based on Sentence Selection with Semantic

Representation, 2014 IEEE 26th International Conference on Tools with Artificial Intelligence

14. Ramesh Nallapati , Bowen Zhou , Cicero dos Santos , Çaglar G ˘ ulçehre ,Bing Xiang, "Abstractive Text Summarization using Sequence-to-sequence RNNs and beyond",arXiv,2016

15. Natural Language Processing in action Understanding, analyzing, and generating text with Python, Hobson Lane Cole Howard Hannes Max Hapke Foreword by Dr. Arwen Griffioen U

BIOGRAPHIES:

Devika Vilasrao Bhise is currently pursuing master's degree program in Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, Maharashtra. Areas of interest and research are Natural language processing, Text processing .

Sneha Narayan Vishwabramha is currently pursuing master's degree program in Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, Maharashtra. Area of interest and research is Natural Language processing with Deep learning