IFERP

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

Text Summarization of Data by Using the Recursive Iteration Technique

[1] Vijayalaxmi M H, [2] Rajeshwari R
[1] PG Scholar, Dept of Computer Science and Engineering, Rao Bahadur Y. Mahabaleshwarappa Engineering College Bellary,
[2] Assistant Professor of Computer Science and Engineering,

Abstract: - **Neural arrangement to-grouping models have given a reasonable new way to deal with abstractive content synopsis (which means they are not limited to just choosing what's more, reworking sections from the first content). Be that as it may, these models have two deficiencies: they are at risk to imitate accurate points of interest erroneously, and they tend to rehash themselves. In this work we propose a Recursive iteration technique in which accuracy of the summarization of the data can be increased. By applying the multiple level of summarization without missing of any important data present in the document the results will be achieved. Here we can apply our models in the news articles which need to be summarized.**

Key words— **Data, Models, Iteration, Technique, Summarization.**

I. INTRODUCTION

Summarization is text that is produced from one or more texts that expresses important information in original text and in shorter form. In this new era, Internet has brought a vast amount of online information. So, every time when someone searches for something on Internet, the response is obtained with lots of different webpages with the more information, which is impossible for a person to read completely. Although many attempts were made to generate the summaries, in recent the field of text summarization has experienced an exponential growth due to emergence of new technologies and algorithms. Generating a condensed version of summary where as preserving its meaning is known as text summarization. Tackling this task is a crucial step towards natural language understanding. Text summarization systems may be loosely classified into two types [2].

1. Extractive model: This model generates summaries by cropping vital segments from the first text and swings them along to create to coherent outline.
2. Abstractive model: In this model, it generates the summaries from the scratch while not being forced to reprocess phrases from the first text.

There are two different groups of text summarization: indicative and informative. Indicative summarization only represents the main idea of the text to the user. The typical length of this type of summarization is 5 to 10 percent of the main text. On the other hand, the informative summarization system gives the concise information of the main text. The length of informative summary is 20 to 30 percent of the main text. Summarization methods also can be classified primarily based upon the quantity of supply documents, task specific constraints and use of external resources as shown within Figure1. Summarization is assessed as single document or multi document Primarily based upon the quantity of supply document, in multi document summary overlaps between totally different documents makes task tough. Primarily based upon the external resources accounts can be classified as knowledge-rich or knowledge-poor. Data made summarizer uses external supply external corpus like Wikipedia, Word Net etc.
In question-focused or question directed account outline is made with info relevant to the query. In update account summarizer makes use of current trends for construction of the outline. The aim of the update outline is to spot new items of knowledge from the document

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

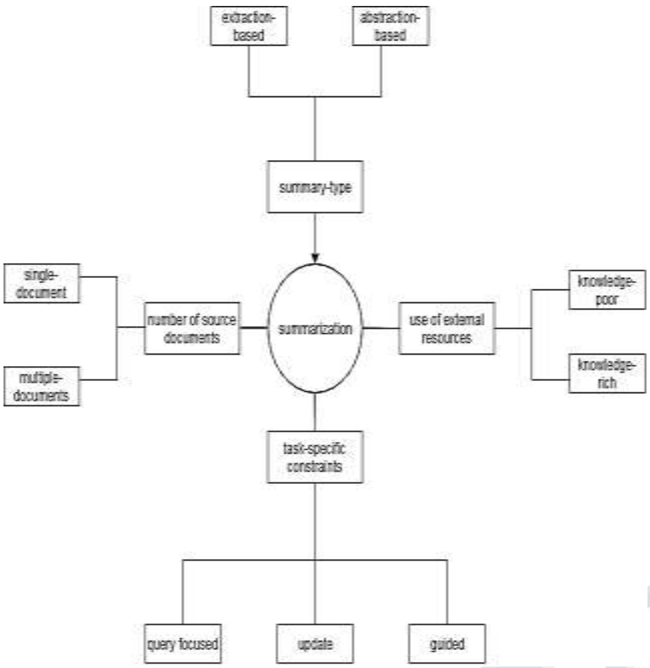


FIGURE 1: Summarization methods

II. RELATED WORK

1. Neural network based encoder-decoder: The neural network based encoder-decoder models are among recent enticing methodologies for try language generation tasks. This model investigates the utility of structural grammar and linguistics info in addition incorporated during a baseline neural attention-based model. We tend to cipher results obtained from associate degree abstract which means illustration programmed employing a changed version of Tree-LSTM. Our planned attention-based encoder-decoder model improves headline generation [1].

2. Neural machine translation model: In this model, unlike the traditional statistical machine translation, the neural machine translation aims at building a single neural network that can be jointly tuned to maximize the translation performance [2]. The models proposes recently for neural machine translation often belongs to a family of encoder-decoders and consists of an encoder that encodes a source sentence into a fixed-length vector from which a decoder generates a translation. The use of a fixed-length vector is a bottleneck in improving the performance of this basic encoder-decoder architecture, and propose to extend this by allowing a model to automatically (soft-) search for parts of a source sentence that are relevant to predicting a target word, without having to form these parts as a hard segment explicitly.

3. Document –level Model: The distributed representation with neural networks has recently shown to be effective in modeling natural languages at fine granularities such as words, phrases, and even sentences [2]. Whether and how such an approach can be extended to help model larger spans of text, e.g., documents, is intriguing, and further investigation would still be desirable. A typical problem of document-level modeling is automatic summarization which aims to model documents in order to generate summaries [3]. The models achieve the state-of-the-art performance, and they significantly benefit from the distraction modeling, particularly when input documents are long.

4. Text-Text generation Technique: A novel technique suggested for text-to-text generation in abstractive summarization. Compared to extraction or previous approaches to sentence fusion, sentence enhancement increases the range of possible summary sentences by allowing the combination of dependency sub trees from any sentence from the source text [4]. It experiments indicate that the approach yields summary sentences that are competitive with a sentence fusion baseline in terms of content quality, but better in terms of grammaticality and that the benefit of sentence enhancement relies crucially on an event co-reference resolution algorithm using distributional semantics.

5. Recurrent neural network (RNN): Abstractive Sentence Summarization generates a shorter version of a given sentence while attempting to preserve its meaning [5]. Introduce a conditional recurrent neural network (RNN) which generates a summary of an input sentence. The conditioning is provided by a novel convolutional attention-based encoder which ensures that the decoder focuses on the appropriate input words at each step of generation. The model relies only on learned features and is easy to train in an end-to-end fashion on large data sets.

6. Sequence-To-Sequence (Seq2Seq) Technique: The important problem in sequence-to-sequence (Seq2Seq) learning referred to as copying, in which certain segments in the input sequence are selectively replicated in the output sequence [6]. A similar phenomenon is observable in human language communication. For example, humans tend to repeat entity names or even long phrases in conversation. The challenge with regard to copying in Seq2Seq is that new machinery is needed to decide when to perform the operation.

IFERP

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 5, Issue 4, April 2018

III. METHODOLOGY

In this method summarization is to merging of the first level summary and second level summary here the first level of summary will be generated based upon the after loading the document the document need to cluster to particular domain after that based upon the keywords generation that is by using Term-frequency that is number of times the particular word repeating in the document based upon that keyword will be generated based upon the keywords the data will be marked in the document and we are extracting the related data and first level of summary will be generated. In the second level of summarization the sentences will send to server. The server will check for the semantic dictionary and if any of the keyword present in the sentences that also present in the semantic dictionary will be continuously updated based upon the recursive iteration. Based upon the keywords matched in semantic dictionary the new summarization will be formed. Finally we are merging the first level and second level summarization.

ALGORITHM: This algorithm is based on the machine learning model and recursively updated with the keywords dictionary on its own to increase point summarization literature population and accuracy. This is continuous process of recursion and for each and every execution of the same input accuracy will be increased based on the semantic dependency dictionary.

PSEUDO CODE:

Input $\longrightarrow \int D_k$ //Input document words

Output $\longrightarrow \sum M_w$ //ranked keywords

Algorithm:

Initialization:

$\int$ $D_k \Leftarrow 0$ //dictionary by words which are most attentive Non-ML

$W_w \Leftarrow 0$ //All document Word vector

n=0 //Count

$M_w \Leftarrow 0$ //marking vector

W_w=FETCH(D_k)

N $\longleftarrow$ size (W_w)

For i: 1 — n

Word=GET (W_w)

For j: 1-size (D_k)

Dict = GET D (J)

If (dict==word)

M_w+= (J)

End

End

II. ARCHITECTURAL DESIGN:

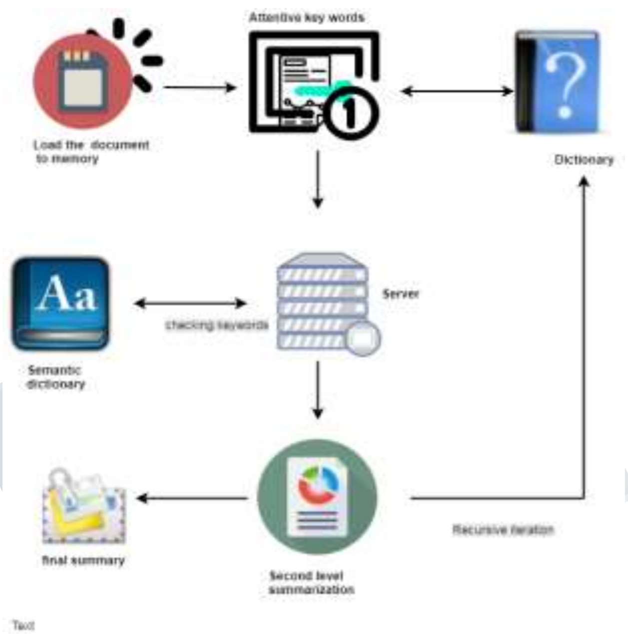


Fig2.Architecture design

The summarization working flow based on the Architectural design as follows below:

1. Load the document: In the first step of the architecture first load the documents into the memory.
2. Attentive keywords: In the second step extract the most attentive keywords based on the keywords extraction. Keywords are marked with the yellow color in the main content. Extract the related sentences based on the keywords.
3. Dependency dictionary: The dependency dictionary is the words where the words will be belong to particular dictionary. There will be dependency between the keywords.
4. Semantic Dictionary: The semantic dictionary where it will checks for the keywords present in the semantic dictionary if any keywords present in the semantic dictionary than in dependency dictionary it continuously updating that keywords to the dependency dictionary.
5. Server: To the server the input will be first level summarization sentences based upon the first level summarization sentences it will check in the semantic dictionary if any keywords present in the semantic

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

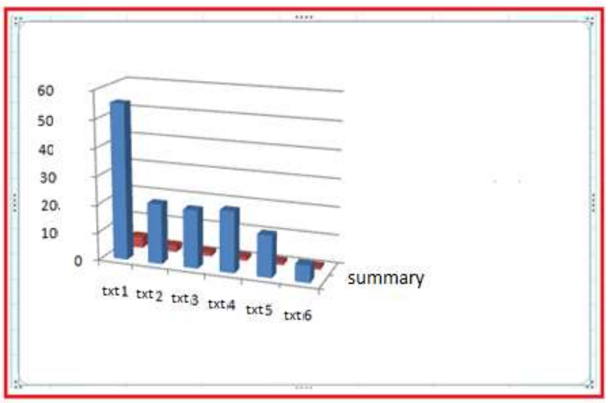
dictionary it will continuously updating keywords to dependency dictionary .

6. Summarization: Based upon the machine learning concept the keywords continuously updating to the dependency dictionary the accuracy will be increased.

7. Final Summarization: After generating the first level and second level summarization finally the merging of the summarization will be done.

IV. EXPERIMENTAL GRAPHS

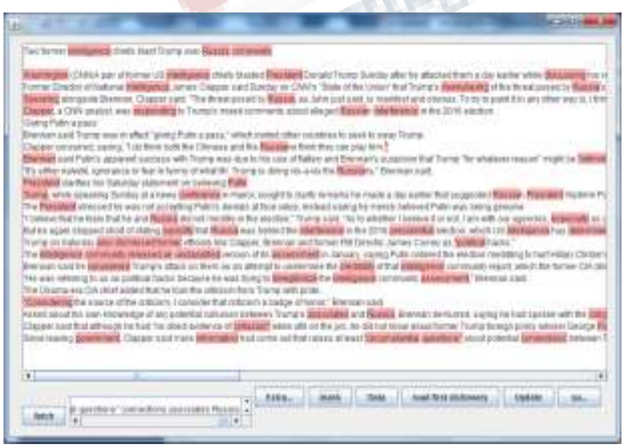
The below graph shows how the 6 different document will be loaded based upon the keywords generated the documents results will be shown.



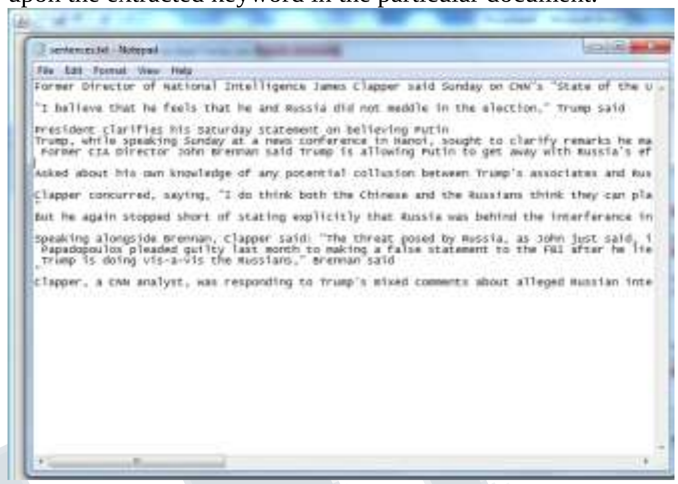
V. RESULTS:

For all experimental, the results are evolved in two levels : First Level and Second Level

First Level Results:

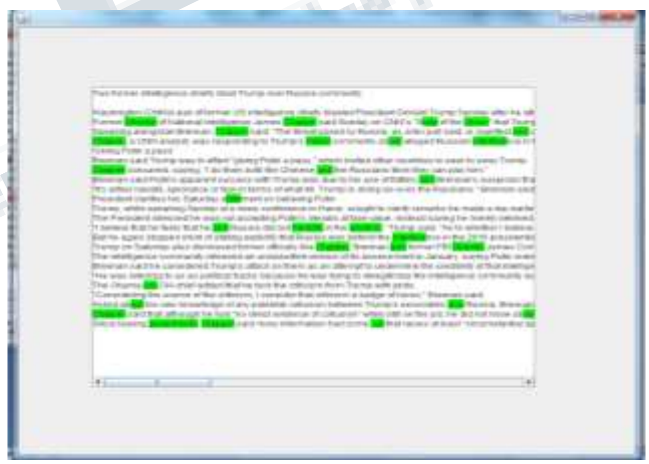


The above screenshots shows the server side Graphical user interface which is comparing the sentences with client sentences and comparing the words with respective dictionary and how the pointers will be marked based upon the extracted keyword in the particular document.



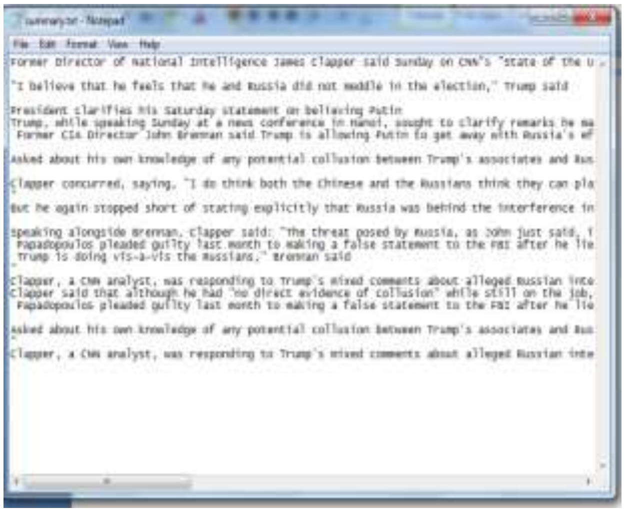
The above screenshots shows the first level of summarization will be generating based upon the extracted pointers.

Second Level Results:



The above screenshots show how the server is comparing the words with the particular semantic dictionary and any of the words present in the sentences also present in the particular dictionary that word will be marking and shows how the new generated keywords will be marked in the client side GUI.

International Journal of Engineering Research in Computer Science and Engineering
(IJERCSE)
Vol 5, Issue 4, April 2018



summary.txt - Notepad

File Edit Format View Help

Former Director of National Intelligence James Clapper said Sunday on CNN's "State of the U

"I believe that he feels that he and Russia did not meddle in the election," Trump said

President clarifies his Saturday statement on believing Putin
Trump, while speaking Sunday at a news conference in Hanoi, sought to clarify remarks he ma
Former CIA Director John Brennan said Trump is allowing Putin to get away with Russia's ef

Asked about his own knowledge of any potential collusion between Trump's associates and Rus

Clapper concurred, saying, "I do think both the Chinese and the Russians think they can pla

But he again stopped short of stating explicitly that Russia was behind the interference in

Speaking alongside Brennan, Clapper said: "The threat posed by Russia, as John just said, t
Papadopoulos pleaded guilty last month to making a false statement to the FBI after he lie
Trump is doing vis-a-vis the Russians," Brennan said

Clapper, a CNN analyst, was responding to Trump's mixed comments about alleged Russian inte
Clapper said that although he had "no direct evidence of collusion" while still on the job,
Papadopoulos pleaded guilty last month to making a false statement to the FBI after he lie

Asked about his own knowledge of any potential collusion between Trump's associates and Rus

Clapper, a CNN analyst, was responding to Trump's mixed comments about alleged Russian inte

The above screenshots shows after merging the first level and second level of summary .final summary will be generated.

VI .CONCLUSION

In this paper we represent a recursive iteration algorithm in which the accurate summary of the data will be generated depending upon the first level and second level of the summarization of data results shown. We applied the model to the news articles where summarization essential.

REFERENCES

[1] Sho Takase "Neural headline generation on abstract meaning representation.

[2] See, A., Liu, P.J. and Manning, C.D., 2017. Get to the point: Summarization with pointer-generator networks. Ar Xiv preprint arXiv: 1704.04368.

[3] Bahdanau, D., Cho, K. and Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[4] Chen, Q., Zhu, X., Ling, Z., Wei, S. and Jiang, H., 2016. Distraction-based neural networks for document summarization. arXiv preprint arXiv:1610.08462.

[5] Cheung, J.C.K. and Penn, G., 2014. Unsupervised sentence enhancement for automatic summarization. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 775-786)

[6] Chopra, S., Auli, M. and Rush, A.M., 2016. Abstractive sentence summarization with attentive recurrent neural networks. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 93-98).

[7] Gu, J., Lu, Z., Li, H. and Li, V.O., 2016. Incorporating copying mechanism in sequence-to-sequence learning. arXiv preprint arXiv:1603.06393.