IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

VOLUME : 5 ISSUE : 1 Print / Issue Publication Date: 13-Jul-2020

ISSN : 2455-2143

DOI : 10.33564/IJEAST.2020.v05i01.086

Indexed In

Google SCHOLAR OPEN ACCESS doi Crossref

WWW.IJEAST.COM

editor@ijeast.com

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 488-492
Published Online May 2020 in IJEAST (http://www.ijeast.com)

EFFICIENT TEXT SUMMARIZER USING POINT TO GENERATOR TECHNIQUE

G. Lasya Sriranga, P. Likitha, B. Meghana
Student, Department Of CSE,
Institute Of Aeronautical Engineering, Dundigal, Telangana, India.
N. Jayanthi
Associate Professor, Department Of CSE,
Institute Of Aeronautical Engineering, Dundigal, Telangana, India.

Abstract —Text summarization is the system of extracting essential facts from the given facts and writing it within the form of precis. Many people locate it tough to read the whole passage of records so that it will gather the essential key factors, so we use textual content summarization to reduce the burden of analyzing big passages. Text summarization is been used in lots of application like business evaluation and marketplace overview. On the entire, by using textual content summarization we get the statistics wished simply by using studying the summary.

Keywords — ***summary, précis, analyze, content extracting.***

I. INTRODUCTION

Text summarization is the method of extracting crucial information from the given statistics and writing it in the form of summary. Lately the want for text summarization has expanded in various paperwork like product evaluation, precis of information, search engine, enterprise analysis. Abstract is described as the precis of information. Since text summarization improved in current years, the want of summary in all fields additionally increased inside the comparable manner. When paper presentation is made abstract has end up obligatory so one can go through the entire statistics in unmarried component. In 1950, automatic summarizer changed into found, which is used to gather all of the crucial sentences from a record and assemble it at one region. The essential intention of this automated summarizer is to lessen the content of files through amassing essential records. Text summarization is divided into two kinds abstractive and extractive summarization. Extractive summarization is the manner of amassing important facts or sentence and mixing it to form a precis without converting the actual which means of the text. Abstractive summarization is the technique of knowledge the source text with the aid of the usage of linguistic technique, studying and analyzing the text. The method entails interruption of text which could certainly make the summary a higher one by means of decreasing the redundancy and maintaining a very good compression rate.

II. LITERATURE SURVEY

Summaries using abstractive techniques are commonly divide d into two types: structured approach and semantic approach.

Structured based approach:

Structured based encodes the most important information from the source text by using extraction rules and templates Techniques used under structured based approach are:

Tree based method: In this method, it uses tree as a dependency to represent the text of a document It uses either a language generator or an algorithm for generation of algorithm

Template based method: In this method, it uses linguistic patterns or extraction rules to extract the information from source text and represent it in the form of template.

Ontology based data: Ontology reduces uncertain data in the source text. It draws relation between each sentence and makes it easy to understand

Rule based method: Documents to be summarized are represented in terms of categories and a list of aspects.

Semantic based approach:

Semantic based approach is used in the framework of natural language generation. This approach is employed by analyzing linguistic data to classify noun phrase and verb phrase.

Semantic-based methodology approaches used include:

Multi modal semantic model: This semantic model captures relationship among concepts. Since it includes salient textual and graphical content it gives excellent abstract summary as an output.

Information item based method: The contents are gathered from the abstract representation and not from source documents. The main strength of this approach is it produces less redundancy and rich information in short summary.

Semantic graph-based method: In this approach it creates a rich semantic graph for the original text resulting in less redundancy and grammatically correct sentences in the summary.

Extractive Summarization Approach: Extractive summarization is the process of collecting important

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 488-492
Published Online May 2020 in IJEAST (http://www.ijeast.com)

information or sentence and combining it to form a summary without changing the actual meaning of the text. Techniques used in extractive based approach are:

Term frequency inverse document frequency method:
The amount of sentences in the document containing
that word is defined as the number of phrases.
Subsequently, these sentence vectors are determined by questi on similarity, and the highest scoring sentences are chosen as part of description.

Cluster based method:
The next element in document (Li) which is the location of the sentence. The last element in the text to which it refers is its resemblance to the first sentence (Fi). Si= W1* Ci + W2* Fi+ W3* Li Where, W1, W2, W3 is overview weight-age. The k-means clustering algorithm is applied.

K nearest neighbor for text summarizing using feature similarity: In this research, we recommend a selected model of KNN (K Nearest Neighbor) wherein the similarity between function vectors is computed thinking about the similarity amongst attributes or features as well as certainly one of values. The undertaking of text summarization is viewed into the binary classification mission where each paragraph or sentence is assessed into the essence or no essence, and in previous works, advanced results are obtained by way of the proposed version inside the textual content category and clustering. In these studies, we define the similarity which considers each attributes and characteristic values, modify the KNN into the model based at the similarity, and use the changed version as the approach to the textual content summarization undertaking. As the blessings from these studies, we may additionally expect the more compact representation of information objects and the higher performance. Therefore, the intention of this research is to implement the text summarization set of rules which represents information items extra compactly and provides the more reliability.

Sentence reduction for automatic text summarization: We gift a singular sentence discount device for mechanically disposing of extraneous terms from sentences which might be extracted from a file for summarization reason. The device makes use of a couple of resources of understanding to decide which terms in an extracted sentence can be removed, which includes syntactic knowledge, context facts, and information computed from a corpus which consists of examples written by means of human experts. Reduction can extensively enhance the conciseness of automatic summaries.

Original text: When it arrives sometime next 12 months in new TV units, the V-chip will supply parents a new and potentially progressive tool to block out packages they do not want their youngsters to see Reduced by humans: The V-chip will provide mother and father a tool to dam out programs they do not want their youngsters to look.

Neural headline generation: It is also an encoder decoder model that uses cutting-edge methodologies to track natural language translation, its syntactic and semantic features of sequence input sentences can be very helpful in generating headline using AMR as it can be proved to be a rooted, guided and acyclic graph encoding the meaning of the sentence. This includes graph nodes representing concepts and directed edges representing the relationship between the nodes. Using this AMR encoder model, the benchmark data showed that standard automated evaluation measures of headline generation tasks have been successfully improved, ROUGE-1, ROUGE-
2 and ROUGEL these findings provide empirical evidence tha t syntactic and semantine information obtained as a result of a n automated parser will help improve the approach of the neur al encoder-decoder in NLG tasks.

III. IMPLEMENTATION METHODOLOGY

Automatic text summarization is producing of summary which has meaning. The input is a collection of reviews or texts and the output may be a quick summary.

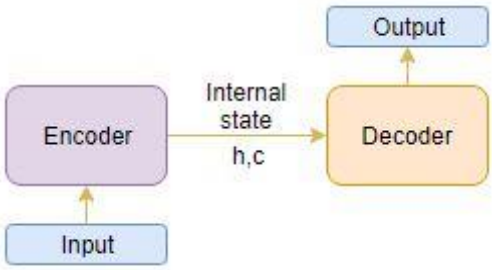

Fig 1. Architecture of Encoder-Decoder

The Recurrent Neural Networks (RNNs) contains Gated Recurrent Neural Network (GRU) or Long Short-term Memory (LSTM) and encoder, decoder.2 steps we will mount the Encoder-Decoder:
1.Training phase
2.Inference section

Training section:
Here we will a set up the encoder and decoder.

Encoder:
An Encoder is LSTM which reads the input, at every timestep, one word is fed into the encoder
The diagram below illustrates this process:

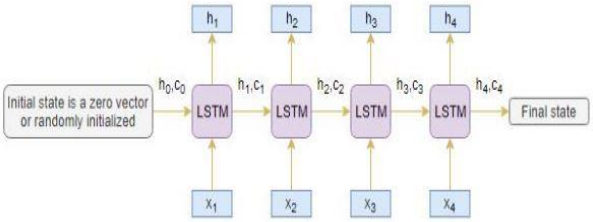

Fig 2. Encoder Model
The decoder is initialized by the hidden state(hi) and mobile

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 488-492
Published Online May 2020 in IJEAST (http://www.ijeast.com)

state(ci) of the ultimate time stage. Know, this is because the encoder and decoder are different units of the LSTM system

Decoder:
The decoder is also a LSTM network, which reads the entire target series term-by-phrase and calculates a one-time offset of the same sequence.
The decoder is trained in anticipating the corresponding sentence in the sequence given the preceding word.

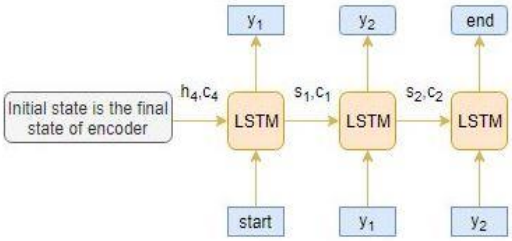

Fig. 3. Decoder model

Beginning and end are special tokens which can be added to the target sequence earlier than those fed into the decoder. Taking a look at the series, the list of priorities is ambiguous. So, we begin to predict the target sequence by passing the first word into the decoder, which could always be the initiating token. And the token finish means the sentence ends.

Inference Phase:
The version is validated after training on new source sequences, for which the aim series is unknown. So, to decode a test series we need to set up the inference architecture:

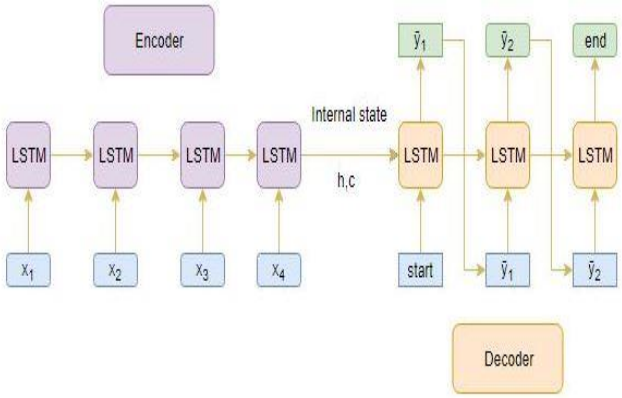

Fig. 4. Inference model

steps to decode the test sequence:

- Encode the input
- Pass a begin token as a variable for input
- Run the decoder for various timesteps
- The output may be a word with maximum occurrence
- Pass the output phrase to the decoder in the inner state time step
- Repeat the steps above until we get a finish token or maximum length of the text is covered

Let's take an instance in which the take a look at collection is given by means of [a1, a2, a3, a4]. The inference process works as below:
Encode the check series into inner country vectors
Observe the following predictions at various time values

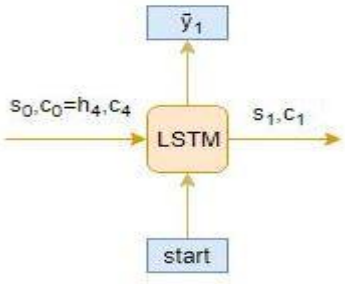

Fig. t=1

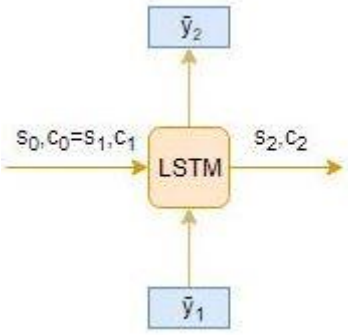

Fig. t=2

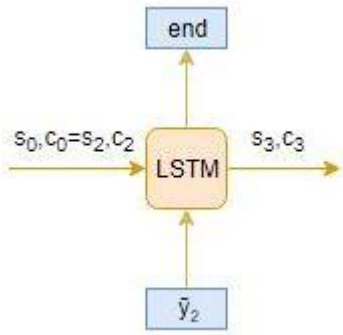

Fig.t=3

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 488-492
Published Online May 2020 in IJEAST (http://www.ijeast.com)

Global Attention:

All of the encoder's hidden states are considered for deriving the background vector in issue:

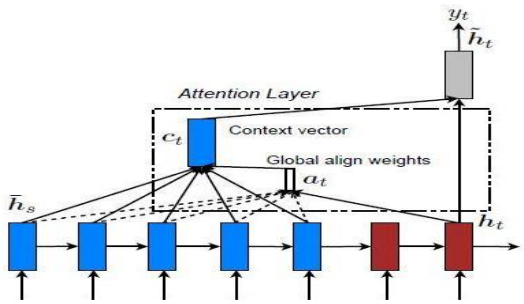

Fig 5. global attention

Local Attention:
Just a few concealed encoder states are known to obtain a vect or for this

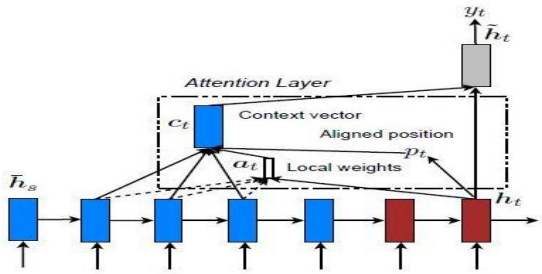

Fig. 6. Local attention

IV. PERFORMANCE ANALYSIS AND RESULTS

Customer reviews may be lengthy and concise at regular intervals. As you might imagine, reading these reviews manually is practically time-eating. This is where it is possible to implement the power of Natural Language Processing to produce an accuracy for long reviews. Our goal right here is to provide a description of the use of the abstraction-based approach for the Amazon Fine Food evaluations.

```
Review: bought several vitality canned dog food products found good quality product look
Summary: good quality dog food

Review: product arrived labeled jumbo salted peanuts peanuts actually small sized unsalt
Summary: not as advertised

Review: confection around centuries light pillowy citrus gelatin nuts case filberts cut
Summary: delight says it all

Review: looking secret ingredient robitussin believe found got addition root beer extrac
Summary: cough medicine

Review: great taffy great price wide assortment yummy taffy delivery quick taffy lover c
Summary: great taffy
```

Fig. 7. Sample reviews of the dataset

Model building
Steps:

- Return Sequences = True: Once set to True, the parameter returns sequences, LSTM produces the secret state and cell nation for each time step
- Initial State: It is used for the initialization of internal LSTM states first.
- Stacked LSTM: Stacked LSTM involves a few layers of LSTM stacked on the top of each other. Which results in a higher representation of the set. Below are a few summaries generated by means of the model:

```
review: really enjoy one tried many top little pricey order sometime future
Original summary: good coffee for strong coffee lovers
Predicted summary:  great coffee

review: ordering bulk repeat shipping option tasty tea great value
Original summary: great value
Predicted summary:  great tea

review: fresh tasty soft go well omaha steak burger store well
Original summary: very fresh and tasty
Predicted summary:  great product

review: son favorite dinner best seems best protein dinners great protein love love
Original summary: favorite
Predicted summary:  best
```

Fig. 8. Summarized reviews

V. CONCLUSION

In this work we have developed a text summarizer using deep learning mechanism on point generator bases, and confirmed that it reduces inaccuracies and repetition. We carried out our model to a new and tough long text dataset, and drastically outperformed the abstractive contemporary result. Our model exhibits many abstractive abilities, however reaching better ranges of abstraction stays an open research question.

VI. REFERENCES

[1] Sho Takase, Jun Suzuki, Naoaki Okazaki, Tsutomu Hirao, and Masaaki Nagata. (2016)- "Neural headline generation on abstract meaning representation. In Empirical Methods in Natural Language Processing."
[2] Jun Suzuki and Masaaki Nagata (2016)- "RNN-based encoder-decoder approach with word frequency estimation." arXiv preprint arXiv:1701.00138
[3] Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, C¸ agla Gulc¸ ehre, and Bing Xiang (2016)- "Abstractive text summarization using sequence-to-sequence RNNs and beyond. In Computational Natural Language Learning."
[4] Ramesh Nallapati, Feifei Zhai, and Bowen Zhou (2017)-. "Summa RuNNer: A recurrent neural network-based sequence

model for extractive summarization of documents. In Association for the Advancement of Artificial Intelligence"
[5] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. (2015)- "Neural machine translation by jointly learning to align and translate. In International Conference on Learning Representations."
[6] Qian Chen, Xiaodan Zhu, Zhenhua Ling, Si Wei, and Hui Jiang. (2016)- "Distraction-based neural networks for modeling documents. In International Joint Conference on Artificial Intelligence."
[7] Jackie Chi Kit Cheung and Gerald Penn (2014)- "Unsupervised sentence enhancement for automatic summarization. In Empirical Methods in Natural Language Processing."
[8] Sumit Chopra, Michael Auli, and Alexander M Rush. (2016)-"Abstractive sentence summarization with attentive recurrent neural networks. In North American Chapter of the Association for Computational Linguistics."
[9] Michael Denkowski and Alon Lavie. (2014)- "Meteor universal: Language specific translation evaluation for any target language. In EACL 2014 Workshop on Statistical Machine Translation."
[10] John Duchi, Elad Hazan, and Yoram Singer. (2011)- "Adaptive sub gradient methods for online learning and stochastic optimization. Journal of Machine Learning Research" 12:2121–2159.

ACKNOWLEDGEMENT

1. **G. Lasya SriRanga** is pursuing B. Tech in the stream of Computer Science and Engineering in Institute of Aeronautical Engineering, Dundigal, Hyderabad, Telangana, India.
2. **P. Likitha** is pursuing B. Tech in the stream of Computer Science and Engineering in Institute of Aeronautical Engineering, Dundigal, Hyderabad, Telangana, India.
3. **B.Meghana** is pursuing B. Tech in the stream of Computer Science and Engineering in Institute of Aeronautical Engineering, Dundigal, Hyderabad, Telangana, India.
4. **N.Jayanthi** is an Assistant Professor in the Department of Computer Science in Institute of Aeronautical Engineering, Dundigal, Hyderabad, Telangana, India.

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

ABOUT IJEAST

International Journal of Engineering Applied Science and Technology (IJEAST) is a peer-reviewed, open access journal that publishes high-quality research papers in the field of Engineering, Applied Science and Technology.

IJEAST aims to provide a platform for researchers, academicians, and professionals to share their innovative ideas, research findings, and practical experiences with the global scientific community.

FOCUS AREAS

- Engineering
- Applied Science
- Technology
- Innovation & Development
- Interdisciplinary Studies

PEER REVIEWED
All submissions are rigorously peer reviewed to ensure quality.

OPEN ACCESS
Free and unrestricted access to research for all.

GLOBAL REACH
Connecting researchers and professionals worldwide.

TIMELY PUBLICATION
We ensure a swift and efficient publication process.

For more information, visit our website
www.ijeast.com

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

editor@ijeast.com
www.ijeast.com
India

2455-2143