International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Query Based Text Summarization

Sindhura Mangalampati, Raveendra Babu Ponnuru

Abstract: ***Now a day's, The advancement of technology implies the large amount of online textual Information, this leads to a need for Text Summarizers which can provide important textual data from large data source like www, into a well-structured document formats. It becomes an alive analysis activity to specify the related data that is retrieved from huge set of documents which are established from various sectors like social media. Text summary is the action of precise the documents by protecting the key concepts of the text. This process can be accomplished via extractive summary and abstractive summary. In this paper, we suggested summarization technique that is to be applied on online information depending upon the User required Document.***

Index Terms: ***Abstractive Summary, Extractive Summary, Information Retrieval, Text Summarizer.***

I. INTRODUCTION

An innovation in Natural Language Processing with machine learning algorithms for language processing was initiated in the late 1980's. This process was implemented due to discouragement of set of corpus syntax which regulates the machine learning methods to process the language. Some of the primitive usage of machine learning algorithms such as decision trees, hard if-then rules are similar to existing hand-written rules. Increasing analysis has focused on statistical methods, which generate easy, probability decisions based on real-valued weights to the features based on input data. Recently, analysis has progressively focused on unsupervised and semi-supervised learning algorithms which are able to learn from data that has a combination of annotated and non-annotated data. In General, this function is more difficult than supervised learning, and it typically produces less accurate results for given input data.

Increasing the availability of online data has demand in the area of automated text summarization within the NLP group [1]. The difficulty has been labelled from different outlook, in various domains and models. Text summary is the process of converting a shorter form of the given larger textual information which contains only vital information and thus, it helps the user to understand the text in a little time period. The main advantage is it reduces user's time in searching the important details in the document [2]. There are mainly two ways to summarize textual documents.

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
Sindhura Mangalampagti, Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India.
Raveendra Babu Ponnuru, Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India.

1. Extractive Way: It involves selecting phrases and sentences from the original text and including it in the final summary.
2. Abstractive Way: The Abstractive method involves achieving exclusively new phrases and sentences to grab the meaning of source document.

II. METHODOLOGY

In this section we describe the each and every step that we perform while we achieve our task to accomplish our objectives. The scenarios are as follows

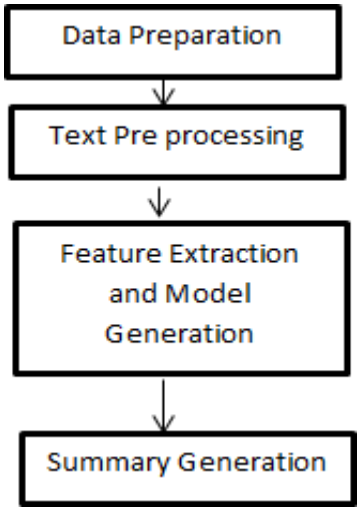

Fig I. Flow chart

1. Data Preparation

The prior step of our project is data preparation. Here data is collection from large amount among multiple sources through a query from online resources [3].

2. Text Pre-processing

The further step after data collection is text pre-processing [4]. As we collect data in the form of text we have to pre-process it. Text pre-processing is a vital part in all text processing functions. It contains some vital parts such as tokenization, Removal of stop words, stemming.

A. Tokenization

It basically refers to splitting a larger body of text into smaller lines, words or even creating words of non-English language. There are two types of tokenization [5]:

Word Tokenization: splitting of given information into meaningful words.

Sentence Tokenization: splitting of information into meaningful sentences [6]. Now, we are using sentence and word tokenizers for feature extraction demonstration.

B. Removal of Stop Words

The words that are appeared repeatedly and regularly which doesn't make any sense [7] like "a", "an", "the" and so on are referred as stop words. So, in this step we can remove those words from the text.

C. Stemming

Stemming is a process that identifies root words, which diminishes the words into their roots [8]. For example, the words "is", "am" and "are" are converted to their root, that is, "be".

3. Feature extraction and Model Generation

It is the process of identifying key features from huge amount of dataset. In this paper, the most explanatory sentences are analyzed and extracted from the given text. The main goal is to choose which sentences are necessary [9][10]. In order to attain, some key features should be extracted from each sentence and there weights are calculated depends on its feature values. Then the important sentences are selected to be present in the summary. Important sentences can be preferred by using word weights and sentence weights using vectors concept [11][12].

A. Word Vector

It is numbered vectors that describe the word meaning. The main advantage of characterizing words as vectors is that they add itself to mathematical operations. It helps to minimize the loss of its predictions. It can be describe words as complex, stable floating point numbers where semantically identical words are charted to proximate points in geometric space. (i.e, it is defined as a row of real valued numbers where each point identifies as a dimension of its word meaning and where semantically identical words have similar vectors is called Word Vector [12]).

B. Sentence Vector

It is defined as an assignment of each sentence to its index and it can be considered as index to any other word [2][3]. The following figure describes the process in detail.

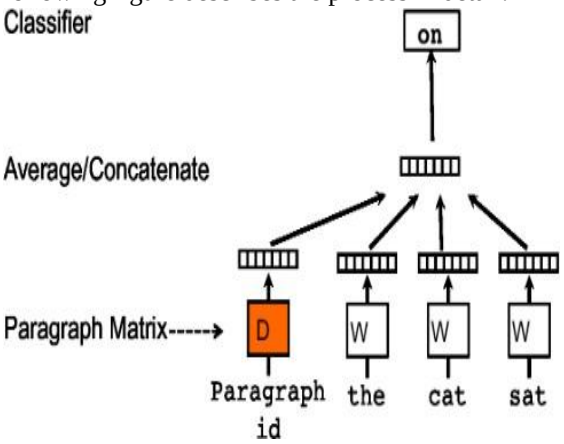

Fig II: Sentence vector

Substantially, every paragraph (or sentence) is graphed to a rare vector, and the combined paragraph and word vectors are helpful to predict the next word.

4. Summary Generation

The further step after the completion of features extraction is summary generation; all over sentences are identified depend on their weights. Topmost sentences are preferred to produce the summary.

III. RESULTS

Let us consider an example of user required query is Text summarization techniques when we perform the same example on proposed model as follows:
Enter a query sentence: Text summarization techniques
Performing Google search for query: "Text summarization Techniques"
Extracted text length from web for your given query: 1098
Calculated length of summary: 274

IV. CONCLUSION

In this paper we study various models to perform summarization. The main protest of text summary is not only balancing query related data word weights and sentence weights can be calculated based on context. In order to improve the scope of generating the summary from set of multiple domains.

REFERENCES

1. P N Varalakshmi K, Jagadish S Kallimani," Survey on Extractive Text Summarization Methods with Multi- Document Datasets,"IEEE conference on Advances in computing,communications and informatics(ICACCI) 2018.
2. Tasuku Kimura, Ryo Tagami, Hisashi Miyamori," Query-focused Summarization Enhanced with Sentence Attention Mechanism," IEEE International Conference on Big Data and Smart Computing (BigComp),2019.
3. Mahsa Afsharizadeh, Hossein Ebrahimpour-Komleh, Ayoub Bagheri, "Query-oriented Text Summarization using Sentence Extraction Technique," 4th International Conference on Web Research (ICWR), 2018.
4. Aditya Jain, Divij Bhatia, Manish K Thakur, "Extractive Text Summarization using Word Vector Embedding," International Conference on Machine learning and Data Science 2017.
5. Yang Wei, Yang Zhizhuo,"Query based Summarization using topic background knowledge," 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2017),2017.
6. Elaheh ShafieiBavani , Mohammad Ebrahimi , Raymond Wong , Fang Chen," A Query-based Summarization Service from Multiple News Sources," IEEE International Conference on Services Computing,2016
7. Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32, ser. ICML'14. JMLR.org, 2014, pp. II–1188–II–1196. [Online]. Available: http://dl.acm.org/citation.cfm?id=3044805.3045025
8. Ghalehtaki, R.A., H. Khotanlou, and M. Esmaeilpour. A combinational method of fuzzy, particle swarm optimization and cellular learning automata for text summarization. in Intelligent Systems (ICIS), 2014 Iranian Conference on. 2014. IEEE.
9. Modaresi, P. and S. Conrad. From Phrases to Keyphrases: An Unsupervised Fuzzy Set Approach to Summarize News Articles. in Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia. 2014. ACM.
10. T. Sri Rama Raju, B. Allarpu, "Text Summarization using Sentence Scoring Method," International Research Journal of Engineering and Technology (IRJET), 2017, Vol. 04, Issue.04
11. S. Km, R.Soumya, "Text summarization using clustering technique and SVM technique", International Journal of Applied Engineering Research, 2015, Vol. 10, pp. 28873-28881

12. R. Nallapati, B. Zhou, C.N. Dos Santos, C. Gulcehre, B. Xiang, "Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond," in Proc. The SIGNLL Conference on Computation Natural Language Learning(CoNLL), 2016, arXiv:1602.06023[cs.CL]

AUTHORS PROFILE

Sindhura Mangalampati, Studying M.Tech, Department of CSE, Velagapudi Ramakrishna Siddhartha Engineering College ,Vijayawada. .

Raveendra Babu Ponnuru, Assistant professor, Department of CSE, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada. he had teaching experience of 7 years. And he published 4 international papers and 2 national conferences.