International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 08, 2020
ISSN: 1475-7192

Automatic audio summarization using Natural Language Processing

[1]Udit Chauhan, [2]Tarun Tiwari, [3]R. Anita

ABSTRACT--*Audio summarization through text summarization is a very important application of Natural Language Processing(NLP). Whenever there is a conversation happening it involves various types of discussions. Important information always gets lost between such jumbled conversations. Consequently, it becomes extremely essential to extract those important key points for future reference. In this paper we aim to implement the automation of the process. The relevant information resulting from the conversation is extracted with the aid of Natural Language Processing techniques by employing text summarization approaches such as Extractive text summarization and Abstractive text summarization.*

Keywords--*Audio Summarization, Natural Language Processing, Automatic Text summarization, Extractive Text summarization, Abstractive Text summarization*

I. INTRODUCTION

In the recent years data has started to grow exponentially. It has been emerging continuously from several forms like text, graphics, audio, video, images, animations etc. Textual data is the most prominent and significant data out of all the above sources. One such source of textual data is text summarization from voice/audio. Almost every type of business is moving towards digitalization. Meetings, conferences, seminars etc. are also no different. It becomes extremely essential to keep an eye over the important information, critical points , agendas etc. As a result these valuable information should be stored safely for future reference. Whenever someone wants to gather information he can get it instantly. But the problem in doing so is that there is bulk of data and storing this data directly without processing is going to incur added costs like maintenance, hardware, software etc. Also it is not efficient. Hence to save time and efforts, summarization of this data becomes important. The summarized data should be meaningful and should be in a clear and concise format .

Text summarization is broadly classified into two parts namely Extractive text summarization and Abstractive text summarization. An extractive summarization method [1] begins with selecting words, sentences and paragraphs from the source. These sentences are given importance based on various criteria. Whether a sentence is important or not , it's decision is based on many linguistic and statistical features which can be at word level or sentence level. Through the method of scoring scheme the importance of a sentence is found out. Then the statements with higher score are selected to generate the summary. Length of the summary is an important aspect and it is decided by the rate of compression.

[1]*SRM Institute of Science and Technology,Chennai, Tamil Nadu,chauhan.udit444@gmail.com*
[2]*SRM Institute of Science and Technology,Chennai, Tamil Nadu,tarun.twr0075@gmail.com*
[3]*SRM Institute of Science and Technology,Chennai, Tamil Nadu,anita.r@ktr.srmuniv.ac.in*

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

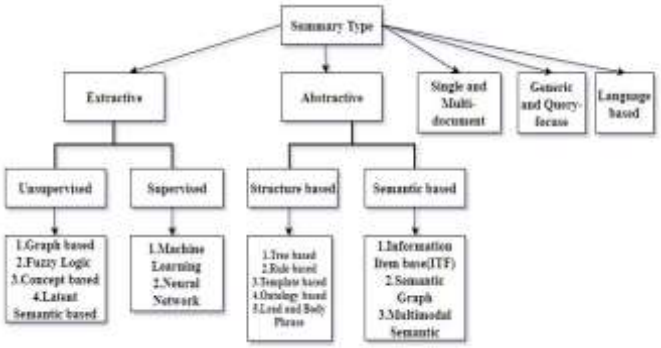
Whereas if we consider abstractive summarization it produces an abstract summary including words and phrases that are different from those occurring in the original document. Hence abstractive summarization is a type of summary that involves concepts and ideas from the original source document but then reproduced in a different form while maintaining the theme and essence of the source. Therefore the concepts of Natural language processing are required here. Hence, it is harder to develop abstractive summaries than extractive summaries.

In this paper the above stated task is achieved by first converting the voice/audio into textual format followed by text summarization and then replying back to the user queries in audio/text format. The paper covers a general overview of the approaches to text summarization and then detailed model along with its implementation.

II. RELATED WORK

Text summarization has its roots originated in late 1950s. Various statistical works were published right from the starting of 1958. Types of summaries are described below in the diagram.

Fig. 1. Types of summaries [2]



Here we will be mainly concentrating the characteristics and methods of extractive text summarization and abstractive text summarization.

A. Extractive text summarization

Several techniques in extractive text summarization methods which revolves around extracting important paragraphs and sentences. It includes both word level and sentence level features which are described below.

As far as word level features are concerned, Content words including important keywords for instance nouns, adjectives, verbs, adverbs etc which should be included in the summary generation. Title words, that are visible primarily in the tagline/title of the text. Biased words also known as domain specific words represent the whole theme of the text document formed by predefined set of words. The quoted words be it singly quoted or doubly quoted and Uppercase words which appear in the document are also good candidates towards summary generation. For example: UNEP(United Nations Environment Programme), IEEE(Institute of Electrical and Electronics Engineers) etc. The positive impact or negative impact words in formation of a sentence are very important to be included in the summary. These words are known as Cue words and are considered of extreme importance. These words can be easily identified as they frequently begin with words like "in conclusion", "the scientist said", "in short", "the writer says" etc. Coming to sentence level features, Length has a major factor in summary generation process. It is evident that longer sentences carry significant information in contrast to shorter length sentences.

Firstly the normalized length of the sentence is found out. It is the ratio of the number of words appearing in the sentence which is under current investigation to the number of words occuring in the maximum length sentence present in the text. Another factor which has a considerable impact towards creation of summary is the location of words. Sentences which are placed generally at the start and end of the paragraph or document are of dominant nature. As a result in most of the summaries it was evident to include the above word and sentence level features as important information to make a better and meaningful summary.

The various approaches to extractive summarization are :

1) *Graph based approach:* Most of the information in any document can be easily described by Graphs. It can be applied to both single-document and multi-documents. Concept similar to HITS algorithm were proposed which efficiently and coherently selects the important sentences. Eigen vector based approach know as LexRank[3] where the sentences represented as graph and the edges represent the similarity. The sentences were then clustered based on their LexRank Scores similar to Page Rank algorithm[4]. Advantages- Effective in areas such as image captions, biomedical documents and newswire, Coherency is improved, Redundant information is identified. Limitations- Doesn't focus on dangling anaphora issue.

2) *Fuzzy logic:* In this model there were four basic components namely Fuzzifier, Inference Engine, Fuzzy Rule base and Deffuzifier. Firstly the source text was cleaned and preprocessed followed by extraction of various features like word level, sentence level etc. As a result, features score were alloted to text items and on the basis of its feature score, sentences were passed to Fuzzifier. Through various rules specified in knowledge base known as rule base, it was passed through inference Engine. Output of inference engine known as Fuzzy output goes through Deffuzifier. This Mechanism leads to the calculation of sentence scores. Based on these sentence scores, summary is generated. To maintain coherency summary is generated in the order of occurrences in the main document. Improvement in coherency was seen.

3) *Concept based Approach:* In this method conceptual vector space model is used to form a rough summarization[5]. For reducing redundancy, degree of semantic similarity is calculated. This approach was efficient as shown by experimental results leading to reliable performance of the system. In this method concepts are extracted from knowledge base such as Wikipedia. Significance of the statement decided by the concepts derived from wikipedia. The steps involved in this process can be summarized as: First extract a concept from external knowledge base. Then build a graph representing the relationships between the derived concepts and sentences. Finally the sentences are sorted on the basis of relative scores and summary is generated. Advantage- similarity measures to reduce redundancy. Limitations-Dangling anaphora , verb references not considered.

4) *Latent Semantic Analysis Method:* The document to be summarized is given as input. After getting input it tries to establish and search numerous patterns and correlations. For instance it may search for words that exist together. Another such pattern may be like the words that are seen only in different sentences[6]. If the frequency of the words that are common is high it stipulates that those sentences are semantically related. There are various methods to find these interrelations between words and sentences, one such method is Singular Value Decomposition[7]. Advantage: Documents were mapped to the same concept space and cluster formation of similar words were seen. External training not required. Limitations: Multiple meaning words weren't handled properly. Gaussian distribution was assumed which is not fitting with some problems.

Received: 21 Dec 2019 | Revised: 18 Jan 2020 | Accepted: 05 Feb 2020

B. Abstractive text summarization

Abstractive text summarization methods includes two approaches namely Structure based and semantic based.

Structure based approach extracts most important information through the text.

1) *Tree based method:* A dependency tree[8] was used to represent the sentences of the source text. Large no of such dependency trees were created and later on these trees were consolidated to form one single tree representing the sentence known as fused sentence. This whole method of conversion of dependency tree to make a summarized sentence is known as tree linearization. The type of parser chosen decides the performance of this method.

2) *Rule based* method*:* In this method the selection criteria and attributes are explicitly defined by the user. On the basis of these rules and patterns content summary is generated. Summaries generated by this method show greater information density. Main drawback of this method is that rules are written manually hence it relies too much on human effort, so process can be tedious.

3) *Template based method:* A template is designed to represent the document. Rules are used to match patterns to template slots. Coherent summaries were created using template based method. Designing of templates is a time taking and difficult task.

4) *Ontology based method:* Entity and their relationships are limited to certain specific domains. These particular domains form the Knowledge base[9]. Category labels were created. It employs fuzzy ontology logic to handle uncertain data and its handling capacity was better than any other previous models. Experiments showed that SVM classifier trained with ontology based method gave better ROGUE scores than standard feature based trained models.

5) *Lead and Body Phrase Method:* This is phrase based method. In this method the sentences which have significant information and good in length were rephrased by continuously inserting and substituting the phrase. Good for revision of a lead sentence which is semantically correct. But it consists of lots of repetition and it's most of the time elapsed on focusing mainly on rewriting techniques.

Semantic based approach: It deals with the linguistic data. It focuses on noun and verb phrases.

1) *Information item based Method*: It results is a well defined information with minimal redundancy. In this approach the summary was generated not by selecting the sentences but through abstract representation of the source text. Subject, Verb, Object were extracted. It's main disadvantage was that it was unable to create meaningful and grammatical sentences.

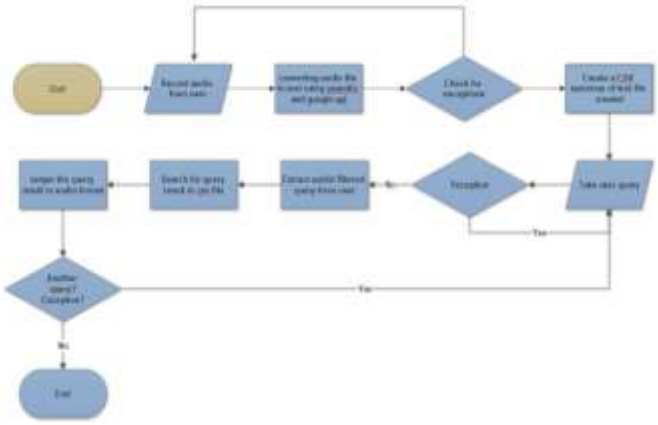
2) *Semantic graph based method*: In this method the words and their relations were represented through semantic graphs. Nouns and verbs forms the node and edges formed the relationships. Intermediate representation using Lexical Chains were proposed. Less redundant and meaningful sentences were created.

3) *Multimodal* semantic *model:* This model emphasize on using the features of more than model. For example consider a 3 stage model. In first stage ontology method used to build the semantic model. Second stage may use Information density matrix for rating a concept, and in last stage graph based model can be used to interrelate concepts to generate summary of high quality . It produces abstract summary with total coverage because it uses textual and graphical features.

III. PROPOSED WORK

In this section we have illustrated, how the architecture shown below is implemented for audio based text summarization using extractive summarization in python. The architecture diagram "Fig. 1" consists of five modules "Fig. 2" which are described as below:

Fig. 2 Architecture diagram automatic audio summarizer



1) Raw Input: Initially the conversation is recorded from sources using input devices like microphone, smartphone etc. The recorded audio should be of good quality. To enhance efficiency of recorded audio extra noise filtering hardware can be used at the input end.

2)*Convertor:* The recorded audio file is converted to text using pyaudio of Python. If there is any exception then control is passed to the exception handler indicating improper conversion of audio file. A CSV file is created from the generated text file in the previous module.

3)*Query Input:* For the first time, user query is taken as input. The query should be meaningful and hence it is checked for exceptions. Extracting useful filtered query from user. Example includes Tell me the summary of the Indian Air Force meeting dated 27th of February. If there is any match related to Indian Air Force meeting on the given date and time, then it will proceed further otherwise will throw an exception saying no such recording exist.

4)Processing: Firstly the text file generated from recorded audio is preprocessed to remove unwanted and unnecessary items/words, which has no impact towards the generation of the summary. Then using the Text Rank Algorithm the important sentences are found. Then using the method of scoring the sentences and phrases , these sentences are aggregated again to form a paragraph or report.

5)*Output* Module*:* The result of the query is conveyed back to the user using the proper output devices like speakers etc.

Example: Let us assume the voice has been recorded by using microphone. It is converted to textual form using pyaudio library of python. Let the converted text be "Wow ... loved this meeting.". Firstly the sentence is converted to lowercase letters, as there is no significance of uppercase and lowercase letters in summary generation. So the sentence will become as "wow ... loved this meeting.", followed by removal of three dots. "wow loved this meeting." Now removing the non-significant words which have nothing to contribute summary like "this", "that", "a", "the" etc. These are also known as Stopwords removal process. Initially it will be in list form with words such

as ['wow', 'loved', 'this', 'meeting']. After Stopwords removal the list will become as ['wow', 'loved', 'meeting']. Now different form of the verbs are reduced to root verb. For example "loved, loving, loves" all will be reduced to "love". It is done because for every word there is one column created in the sparse matrix. Hence more the words, more will be the entries leading to more evaluation time. Now the list will become ['wow', 'love', 'meeting']. For predicting the outcome of this statement whether this list is converted back to string form "wow love meeting". Now it will be easier to predict the importance of this sentence than the actual form.

Algorithms used:

1)Text Rank Algorithm:

i. Read the text-input converted from voice/audio recording.
ii. Split the text into sentences and store in a python list.
iii. Start cleaning the text.
iv. Create a similarity matrix with probability of each sentence.
v. Define the threshold for sentences.

2)Heapq-Nlargest:

Find out the top useful sentences using sentence scores and sorting them in the order.

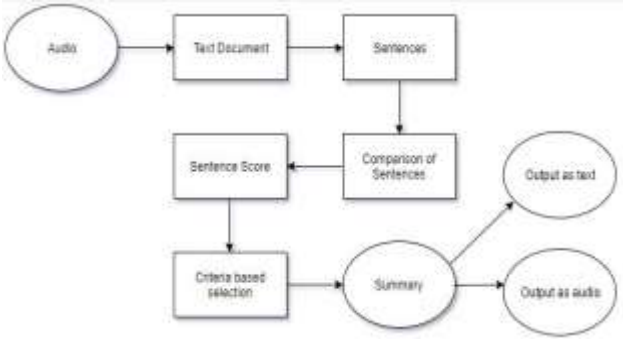
3) TextBlob

Extracting the few random nouns and give only user relevant information about topic.

IV. RESULTS AND DISCUSSIONS

In this section we have illustrated, how the proposed architecture generates the summary.

Fig.3 Module diagram depicting implementation model



The pre-recorded audio/voice were given as input to the model and then it's predicted output is matched with actual human generated summaries. If we somehow eliminate the noises at the source end and cleaned input is taken then it generated summaries with good accuracy. For performance measures , Recall-Oriented Understudy for Gisting Evaluation (ROGUE) is used. It compares the automatic system summary generated by machines against a set of reference summaries typically human-produced. Let the *system generated* summary be: “The IOT conference is postponed.” and the reference summary used be “The IOT conference is postponed to the next month.” For quantitative evaluation , overlapping method of Precision and Recall is used. Recall here signifies how much the system generated summary is able to capture or recover the reference summary. Recall in our model is given by total no of overlapping words divided by the total words in the reference summary used. Hence in our

example, Recall is 5/9 which equals 0.56. Whereas Precision means how much the system summary generated by the model was needed or relevant. Precision in our model is given by the total no of overlapping words divided by the total words in system summary. Hence in our example Precision is 5/5 which equals 1.00. But if the system summary has been like "The IOT Conference is postponed due to some unavoidable reasons." , then Precision would have been 5/10 which equals 0.50. Similarly by using the proper Precision and Recall values, the F-measure can be found out which is defined as twice the product of Precision and Recall wholly divided by the summation of Precision and Recall.

V. CONCLUSION AND FUTURE RESEARCH

In the research work proposed, we have studied audio summarization through text summarization techniques and have presented a model using natural language processing. Huge volumes of research work has already been done on extractive text summarization but because of the complex nature of abstractive text summarization , very less research work has been done on it as compared to extractive approach. In our proposal, we have shown a model with audio to text summarization and its implementation. In future we would like to continue our research with other techniques for aiming better results. Furthermore, this model can also be extended to different regional languages and concepts of deep learning and neural networks can be used to train and test the model.

VI. ACKNOWLEDGEMENTS

We would like to thank all the professors of department of computer science and engineering who directly or indirectly guided us in our research work. We would like to express special gratitude to Dr. Subalalitha C.N , Associate Professor for sharing her immense knowledge and guiding us in overcoming numerous obstacles and encouraging us to get results of better quality.

REFERENCES

1. Gupta V, Lehal GS. A survey of text summarization extractive techniques. Journal of emerging technologies in web intelligence. 2010 Aug 20;2(3):258-68.
2. Chauhan U, Tiwari T. Automatic Text Summarization and it's Methods-a Survey.
3. Erkan G, Radev DR. Lexrank: Graph-based lexical centrality as salience in text summarization. Journal of artificial intelligence research. 2004 Dec 1;22:457-79.
4. S. M. R .. W. T. L., Brin, The page rank citation ranking: Bringing order to the web, Technical report, Stanford University, Stanford, CA., Tech. Rep., (1998).
5. Wang M, Wang X, Xu C. An approach to concept-obtained text summarization. InIEEE International Symposium on Communications and Information Technology, 2005. ISCIT 2005. 2005 Oct 12 (Vol. 2, pp. 1337-1340). IEEE.
6. Gong Y, Liu X. Generic text summarization using relevance measure and latent semantic analysis. InProceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval 2001 Sep 1 (pp. 19-25). ACM.

7. Ozsoy MG, Alpaslan FN, Cicekli I. Text summarization using latent semantic analysis. Journal of Information Science. 2011 Aug;37(4):405-17.
8. Hirao T, Nishino M, Yoshida Y, Suzuki J, Yasuda N, Nagata M. Summarizing a document by trimming the discourse tree. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP). 2015 Nov 1;23(11):2081-92.
9. Ramezani M, Feizi-Derakhshi MR. Ontology-Based Automatic Text Summarization Using FarsNet. Advances in Computer Science: an International Journal. 2015 Mar 31;4(2):88-96.