International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET22928

Automatic Generation of Minutes of Meetings

Anuj Pandya, Prof. Namrata Gawande
Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, Maharashtra, India

Article Info
Volume 9, Issue 2
Page Number : 93-99

Publication Issue :
March-April-2022

Article History
Accepted : 15 March 2022
Published: 21 March 2022

ABSTRACT

This paper describes the process for automatic generation of minutes of meetings using Machine Learning algorithms and Natural Language Processing techniques. Minutes of meetings are a record which are used to keep official summaries of all meetings conducted within a company or organization. Automatic generation of minutes of meeting is a challenging issue and has gathered a huge amount of interest over the last few years due to its applications. Initially, we study previous research papers to understand existing techniques used for the purpose. Techniques such as AMBOC Model, BART Summarizer, HMNet Model, MSCG are employed for detecting useful and informative action items from audio files. Then we explore Machine Learning models such as SVM, HMM which are clubbed with the majority of methods for classification and summarization of the words given by above mentioned models to generate an informative summary for the user.

Keywords : Natural Language Processing (NLP), Machine Learning (ML), Minutes of Meeting,Support Vector Machine (SVM), Hidden Markov Model(HMM), Speech-to-Text.

I. INTRODUCTION

In present times, since the Covid-19 pandemic majority of work all over the world has been done online. In dealing with work online the conduction of online meetings has been and will always be the most crucial factor as it facilitates communication amongst the people whether its any company, educational institute or any other organization. Since people's working hours have significantly increased while working from home, the number of meetings conducted daily have also increased. With this, keeping track of all the daily developments and proceedings in various meetings have become a tedious and time consuming task.

Minutes of meetings are an essential record which helps with this issue. It helps in keeping track of the daily developments and important points discussed in the meetings conducted in an organization. Generally these minutes of meetings are made manually by people but its automatic generation has a wide range of applications and can help in increasing the productivity of people by saving crucial time spent in manual creation of minutes of meetings.

This automatic generation of minutes of meetings is proposed by using Machine Learning models/algorithms and Natural Language Processing techniques. The proposed method involves dividing the whole process into three sub-processes or modules each of them dealing with an important aspect of the generation. The whole process is divided into speech-to-text conversion, text classification and text summarization modules to provide a summary or minutes of meetings as an output.

II. LITERATURE SURVEY

In order to learn in detail about this domain and the previous research done in this field, several studies and research papers were referred to. Details of some of the referred papers are mentioned below with publication time of the paper ranging from newest to oldest.

Megha Manuel et. al in 2021,[1] mentions creating an Automated Minute Book Creation (AMBOC) model using machine learning to extract key information from important discussions in a meeting. Although the model shows good accuracy and can differentiate between male and female speakers, it does not perform satisfactorily while using other languages except English.

Jia Jin Koay et. al in 2021,[2] apply a sliding window approach combined with the BART summarizer to obtain summaries for evaluation. Its advantages are that it does not require annotated data and can be scaled to various domains for usage. But it shows low accuracy in some cases such as detection of highly relevant utterances.

Chenguang Zhu et. al in 2020,[3] mention using an Hierarchical Meeting Summarization Network Model or HMNet Model which is based on an encoder-decoder transformer structure. It shows good performance in terms of both automated and human evaluation metrics but cannot properly cover detailed items during long meeting transcripts.

Anna Nedoluzhko et. al in 2019,[4] analyze required tools and datasets for automatic minuting of meetings and lay out a topology of types of methods,meetings,etc. But they fail at creating a final use-able model for the purpose.

Guokan Shang et. al in 2018,[5] propose using a framework based on Multi Sentence Compression Graph (MSCG). The method produces well framed summaries when compared to human written transcripts but not all outputs are usable due to lack of coherence among several entities.

Siddhartha Banerjee et. al in 2016,[6] propose a method where supervised learning is used to separate various topic segments, and then identify the important utterances. The important utterances are then combined together. The best sub-graph is then obtained by integer linear programming (ILP) which is selected as the final output

Tatsuro Oya et. al in 2014,[7] proposed using a novel multi sentence fusion algorithm to generate summary templates using lexico-semantic information. The template is then selected using relation between human evaluated summaries and their source transcripts.

Yashar Mehdad et. al in 2013,[8] propose using an Entailment Graph to identify semantic relations between sentences and discover important information from a cluster of sentences. This information is then used to build a word graph model which summarizes meeting transcripts. The model outperforms all baselines in ROGUE-1 score but ranks 3rd in ROGUE-2.

Lu Wang et. al in 2012,[9] use an novel unsupervised framework based on existing in-domain relation learner. Clusters of Decision Related Dialogue Acts (DRDAs) are taken as input by the system to generate abstractive summaries. The model outperforms many baseline models and generates competitive summaries in comparison to supervised frameworks.

Lu Wang et. al in 2012,[10] proposed using DomSum, a token-level decision summarization framework used to extract words from Dominant topics to form Summaries. It takes clusters of DRDAs with topic and word distribution as input and the output is a set of topic-coherent summary worthy words which can be directly used as summaries or can further be evaluated.

Justin Jian Zhang et. al in 2011,[11] used a method of automatically generating Parliamentary meeting minutes with one step chunking, parsing and extractive summarization system along with a single Conditional Random Field classifier. The proposed system outperforms two-step models for the same purpose.

Lu Wang et. al in 2011,[12] explored methods of producing summaries of spoken meetings at both token-level and DA-level. It is shown that clustering DRDAs before identification improves the performance of the system.

Gabriel Murray et. al in 2010,[13] proposed a system for generating abstracts of meeting conversations. The summarizer first maps sentences and identifies message patterns. Then the most informative messages are selected using ILP optimization to generate abstract summaries.

William Morgan et. al in 2009,[14] show that several classes of features are useful for annotation of action items from meeting audios. A maximum entropy model to automatically detect action item related utterances from multi-party audio meeting recordings. The effects of various factors such as lexical, temporal, syntactic features on system performance are investigated.

Gabriel Murray et. al in 2008,[15] uses features such as prosodic, lexical, structural and speaker-related to identify most informative words from each meeting. Then meta comments are identified to create more informative summaries which provide an increased level of abstraction. Thus meta comments help in improving summarization performance of the system.

Raquel Fernandez et. al in 2008,[16] propose two approaches for extracting information. The first approach uses an open domain semantic parser to identify candidate phrases for decision summaries and then employs ML techniques such as SVM and HMM for classification. The second approach uses categorical and sequential classifiers that use syntactic and semantic features to identify relevant words and sentences.

Raquel Fernandez et. al in 2005,[17] use approaches using prosodic and lexical features with MMR and LSA approaches for summarization. Also how the summarization results might get affected when carried out on ASR output as opposed to manual transcripts is studied.

Anne Hendrik Buist et. al in 2004,[18] propose using a Maximum Entropy based extractive summarization system which uses a combination of 15 features thus improving baseline systems by selecting all utterances longer than 10 words. This method improves the overall performance of the system by 20%.

Sadaoki Furui et. al in 2004,[19] present methods for speech-to-text and speech-to-speech summarization using speech unit extraction and concatenation. A two-stage summarization technique is used. It includes extraction of sentences and word-based sentence compaction. Sentences, words, and between-filler units are taken as units to be identified from the original data.

Chiori Hori et. al in 2003,[20] proposed a method in which a set of words is extracted and then joined to generate better summaries. This process is achieved using Dynamic Programming (DP).

III. Algorithmic Study

Fig1. Various MoM Generation Techniques

After studying the previous studies in this field, the following models and frameworks have been identified as major techniques employed so far in this direction over the years.

AMBOC model or Automated Minute Book Creation model is used in [1] which further is partitioned into 3 smaller models of speech recognition, speaker verification and text summarization respectively.It supports 119 languages including English, Hindi, Arabic, etc and uses Google API for better text to speech and text summarization results. It shows an overall accuracy of 83.33.

HMnet Model or Hierarchical Meeting summarization Network model used in [3] achieves 11.62 higher ROUGE-1, 2.60 higher ROUGE-2 and 6.66 higher ROUGE-SU4 points than previous best models. It employs a two-level hierarchical structure for cases like long meeting transcripts, and uses a role vector to represent each participant in the meeting.

The BART model proposed in [2] is a pre-trained model which denoises and filters corrupted input text. The encoder learns meaningful representations and the decoder reconstructs the original text using the representations. This method is then combined with a sliding window approach for better evaluation of transcripts.

MSCG or Multi-Sentence Compression Graph used in [5] is a fully unsupervised approach for generating a short, self-sufficient sentence from a cluster of related, overlapping sentences. This is combined with preprocessing text and keyword extraction to form a unified end-to-end summarization framework which outperforms most of the baselines in ROUGE-1 2 scores.

Entailment Graph in [8] is used to identify semantic relations between sentences and thus can be used to discover important information from a cluster of sentences. This information is then used to build a word graph model which summarizes meeting transcripts. The model outperforms all baselines in ROGUE-1 score but ranks 3rd in ROGUE-2.

DomSum is a token-level decision summarization framework used in [10] to extract words from Dominant topics to form Summaries. It takes clusters of DRDAs with topic and word distribution as input and the output is a cluster of summary worthy words which can be directly used as summaries or can further be evaluated.

4. Proposed Method and Implementation

Fig1. Block Diagram of proposed method

The overall process is divided into three modules of speech-to-text conversion, text classification and text summarization. The implementation of the proposed method is done using Python language and its modules and libraries.

1. Speech-to-Text Conversion module

The first module of speech-to-text conversion involves converting the input audio file into plain text. The output plain text is saved into a .txt file which is used further in the next modules and stages of the process.

First the speech-to-text conversion is done using the Speech Recognition module which is a library used for converting audio to text. It uses Google API for this purpose and supports a number of engines and APIs. First the recognizer is initialized and the audio file to be translated is read. The audio file should be preferably in .wav format.

Then the audio file is fed to the Google API recognizer for conversion. Additionally, exception handling is used to avoid issues in conversion due to too much noise and too long length of audio file. A word to word conversion of the speech is achieved as is given as output which is then stored in a .txt file for further processing.

2. Text Classification module

Text classification is a very popular and widely used Machine Learning application. Here the use of text classification is to distinguish action items i.e classify summary worthy and non-summary worthy words. Since the efficiency of the system is an important factor which in this case is related to the informativeness of the final summary generated for the user, text classification is a very important part of the process. The two popular text classifiers used are Support Vector Machine or SVM and Hidden Markov Model or HMM.

Here we use SVM for classification purposes. It is a supervised machine learning model which classifies items by creating a hyperplane in an N-dimensional space depending on the number of features. The .txt file which contains the speech to text converted output of the previous module is given as input here. The implementation involves usage of various python modules and libraries such as pandas, numpy, sklearn and matplotlib. The model is trained, tested and fitted to get the more informative summary-worthy words from the initial bag of words. These classified words are then further processed in the next module.

3. Text Summarization module

The next and final module in the process in the text summarization module. Similar to text classification, text summarization is also a very popular and widely used application of the Machine Learning domain. This is required in the process to summarize the classified words and hence provide a final optimized output to the user.

Here Latent Semantic Analysis or LSA and Maximal Marginal Relevance or MMR are popular Machine Learning models used for summarization. The implementation involves using the Natural Language Toolkit or nltk library in python. The words are tokenized and a dictionary is used to find frequently occurring words. This is further used to provide the final summary to the user.

IV. Conclusion

In this ever evolving and pacing world, with the increase in online tasks and work, minutes of meetings have become a crucial resource which is used to monitor and follow the day-to-day activities and developments in an organization. Hence a method is proposed to automatically generate summaries of audio recordings to save time and avoid manual creation of the same. The whole process is divided into three modules, namely speech-to-text module, text classification module and text summarization module.

After all three modules are done, we obtain a text summary of the initial audio file or recording given as an input. Similarly, a meeting recording can be used to generate a text summary about it which can be used for various purposes by various users. We have used general algorithms such as SVM which are easy to implement and have good accuracy along with a number of python modules and libraries to achieve this. These generated summaries are multipurpose and can come handy even in cases where someone misses the meeting and needs to be updated with the developments or can be used in other places such as educational institutes, seminars, conferences, etc to get a gist of the whole event or talk or lecture.

Still, the efficiency of the method depends on how informative and useful the generated summary is. This aspect may change from user to user depending on the situation or case in which the method is used. A big future scope remains in this domain to be explored regarding such issues and additional advanced problems such as summarization in case of languages other than english and their efficiency.

V. REFERENCES

[1]. Manuel, Megha; Menon, Amritha S; Kallivayalil, Anna; Isaac, Suzana; K.S, Dr. Lakshmi”

Automated Generation of Meeting Minutes Using Deep Learning Techniques" IJCDS July 2021

[2]. Koay, Jia Roustai, Alex Dai, Xiaojin Liu, Fei. (2021). A Sliding-Window Approach to Automatic Creation of Meeting Minutes.2021 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop. 68-75. 10.18653/v1/2021.naacl-srw.10.

[3]. Zhu, Chenguang Xu, Ruochen Zeng, Michael Huang, Xuedong. (2020). A Hierarchical Network for Abstractive Meeting Summarization with Cross-Domain Pretraining. Findings of the Association for Computational Linguistics: EMNLP 2020. 194-203. 10.18653/v1/2020.findings-emnlp.19.

[4]. Nedoluzhko, Anna and Ondrej Bojar. "Towards Automatic Minuting of the Meetings." ITAT (2019). 19th Conference Information Technologies - Applications and Theory (ITAT 2019).

[5]. Shang, Guokan Ding, Wensi Zhang, Zekun Tixier, Antoine Meladianos, Polykarpos Vazirgiannis,Michalis LORRE, Jean-Pierre. (2018). Unsupervised Abstractive Meeting Summarization With Multi-Sentence Compression and Budgeted Submodular Maximization. 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)664-674.10.18653/v1/P18-1062.

[6]. Banerjee, Siddhartha & Mitra, Prasenjit & Sugiyama, Kazunari. (2016). Abstractive Meeting Summarization Using Dependency Graph Fusion.The 24th International World Wide Web Conference (WWW 2015 Companion).

[7]. Oya, T., Mehdad, Y., Carenini, G., Ng, R. (2014). A Template-based Abstractive Meeting Summarization: Leveraging Summary and Source Text Relationships. INLG. 8th International Natural Language Generation Conference (INLG).

[8]. Mehdad, Yashar Carenini, Giuseppe Tompa, Frank Ng, Raymond. (2013). Abstractive Meeting Summarization with Entailment and Fusion. 136-146. 14th European Workshop on Natural Language Generation.

[9]. Wang, Lu Cardie, Claire. (2012). Focused Meeting Summarization via Unsupervised Relation Extraction. SIGDIAL Conference 5 July 2012.

[10]. Wang, Lu Cardie, Claire. (2012). Unsupervised Topic Modeling Approaches to Decision Summarization in Spoken Meetings. 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue.

[11]. Zhang, Justin Jian Fung, Pascale Chan, Ricky. (2011). Automatic minute generation for parliamentary speech using conditional random fields. Acoustics, Speech, and Signal Processing,1988.ICASSP-88.,1988InternationalConferenceon.5536-5539.10.1109/ICASSP.2011.5947613.

[12]. Wang, Lu Cardie, Claire. (2011). Summarizing decisions in spoken meetings. Workshop on Automatic Summarization for Different Genres, Media, and Languages 16-24.

[13]. Murray, Gabriel Carenini, Giuseppe Ng, Raymond. (2010). Generating and Validating Abstracts of Meeting Conversations: a User Study. 6th International Natural Language Generation Conference.

[14]. Morgan, William Chang, Pi-Chuan Gupta, Surabhi Brenier, Jason. (2009). Automatically Detecting action items in audio meeting recordings. SIGDIAL Workshop 15 July 2009. 96-103.10.3115/1654595.1654614

[15]. Murray, Gabriel Renals, Steve. (2008). Meta Comments for Summarizing Meeting Speech. INTERSPEECH 2005. 5237. 236-247. 10.1007/978-3-540-85853-9 22.

[16]. Fernandez, Raquel Frampton, Matthew Dowding, John Adukuzhiyil, Anish Ehlen, Patrick Peters, Stanley. (2008). Identifying relevant phrases to summarize decisions in spoken meetings. INTERSPEECH 2008. 78-81

[17]. Murray, Gabriel Renals, Steve Carletta, Jean. (2005). Extractive summarization of meeting recordings. INTERSPEECH 2005. 593-596

[18]. Buist, Anne Kraaij, Wessel Raaijmakers, Stephan. (2004). Automatic Summarization of Meeting Data: A Feasibility Study. Computational Linguistics in the Netherlands 2004

[19]. S. Furui, T. Kikuchi, Y. Shinnaka and C. Hori, ”Speech-to-text and speech-to-speech summarization of spontaneous speech,” in IEEE Transactions on Speech and Audio Processing, vol.12, no. 4, pp. 401-408, July 2004, doi: 10.1109/TSA.2004.828699.

[20]. Chiori Hori, Sadaoki Furui, Rob Malkin, Hua Yu, and Alex Waibel. 2003. A statistical approach to automatic speech summarization. EURASIP J. Adv. Signal Process 2003 (January 2003), 128–139

Cite this article as :

Anuj Pandya, Prof. Namrata Gawande, " Automatic Generation of Minutes of Meetings, International Journal of Scientific Research in Science, Engineering and Technology(IJSRSET), Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 9, Issue 2, pp.93-99, March-April-2022. Available at doi : https://doi.org/10.32628/IJSRSET22928
Journal URL : https://ijsrset.com/IJSRSET22928