International Journal of Research in Engineering, Science and Management
Volume-2, Issue-4, April-2019
www.ijresm.com | ISSN (Online): 2581-5792


# Secure Multi-Modal Summarization using Machine Learning

Gopish Mundada[1], Piyush Nimonkar[2], Rashmi Kabra[3], Ruchali Sudke[4], N. P. Kulkarni[5]
[1,2,3,4]*Student, Department of Information Technology, Smt. Kashibai Navale College of Engineering, Pune, India*
[5]*Professor, Department of Information Technology, Smt. Kashibai Navale College of Engineering, Pune, India*

***Abstract*: Text summarization aims to condense a source text into a shorter version. Automatic data summarization is part of data mining. The rapid growth in data transmission over the internet makes it necessary to create multi-modal summarization (MMS) from asynchronous data (text, audio, and video). In this work, an MMS method combining the techniques of natural language processing (NLP), speech processing, computer vision and advanced encryption standard (AES) encryption is used to explore the rich information contained in multi-modal data, to improve the quality and security as well the key idea is to bridge and lessen the semantic gaps between multi-modal data. Video is basically composed of audio and visual (image). For audio, speech transcriptions are used. For visual information, the joint representations of image and text are studied using an artificial neural network. Finally, all the multi-modal aspects are considered to generate a textual summary by maximizing the salience, readability, non- redundancy. The summary so generated by text, audio or video is encrypted using AES encryption method (to make it secure) and stored on server, from where the user can retrieve it whenever required by providing the decryption key.**

***Keywords*: AES, Computer vision, Decryption, Encryption, Multimedia, Multi-modal, Natural Language Processing, Summarization.**

## 1. Introduction

Text summarization plays an important role in everyday life and has been studied from past decades. With the evolving information age and the emergence of multimedia technology, multimedia data (including text, image, audio and video) have increased drastically. Multimedia data have greatly changed the way people live and make it difficult for users to obtain important information efficiently. Intuitively, readers can grasp the gist of the event more easily by scanning the image or the video than by only reading document, and thus the multi-modal data will also reduce the difficulty for machine to understand a particular event [6].

Most of the summarization systems or techniques focus only on natural language processing (NLP). Multimedia data is generally and mostly asynchronous in real life, which means no explicit description or information for images and no subtitles for videos is provided. Therefore, MMS faces a major challenge in understanding the semantics of visual information [6]. The proposed system jointly optimizes the quality of the summary with the help of automatic speech recognition (ASR) and computer vision (CV) processing system.

In this work, an MMS system is presented that can provide users with textual summaries to help to acquire the gist of asynchronous multimedia data in a short time without reading documents or watching videos from beginning to end [6].

Due to the rapid growth in the network applications, the security has become crucial factor in the communication and transfer of data over network [5]. Therefore, encryption is used to make data more secure. The different terms regarding the encryption are given as follows:

*Plain text* is the data to be transmitted to the receiver.

*Cipher text* is text produced after encryption.

*Encryption* is the process of converting the data into cipher text.

Decryption is the process of converting the cipher text back into original data.

*Symmetric Encryption:* In symmetric encryption the same key is used for both encryption and decryption of the data.

*Asymmetric Encryption:* In asymmetric encryption different keys are used for encryption and decryption [5].

## 2. Literature survey

The intention of text summarization is to express the content of a document in a condensed form that meets the needs of the user. It isn't possible to read everything or listen to an audio or watch a video from start to end so some form of information condensation is needed.

Multimedia data is generally and mostly asynchronous in real life, which means no explicit description or information for images and no subtitles for videos is provided. And the studies conducted so far deals well with synchronous data; therefore, a system dealing with asynchronous data is required.

The proposed system aims to create the summary from asynchronous data. Further the generated summary is stored onto the server. The summary is first encrypted using AES encryption method for security issues and then further stored onto server. Various methods are available such as DNA (Deoxyribo Nucleic Acid) Cryptography. DNA Cryptography provides multiple security layers, but it requires high tech lab requirements and is complex in nature [4]. Therefore, AES encryption method is used to encrypt the generated summary which provides better results and is feasible.

## 3. Proposed system

### A. Workflow diagram

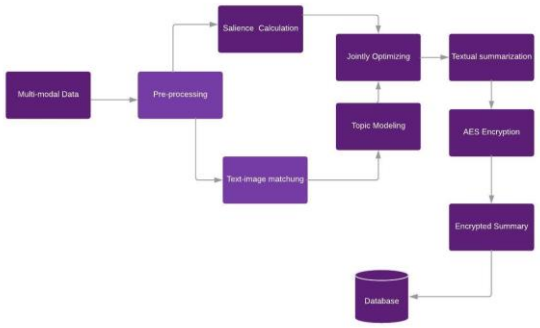


Fig. 1. Work flow of proposed system

Fig. 1 shows the basic work flow of proposed system. Multi-Media data is collected from the source. Collected data is then sent for pre-processing. In the pre-processing stage it gets divided into different modules; depending upon it is audio video or text. First is salience calculating where audio extraction from the video is done by using ASR. Secondly there is text-image matching model where the extraction of text and the image is done. After the text-image matching the extraction of topic modeling is done by using the information gathered by text-image matching model. In this way, the jointly optimization of the audio, image and text is combined together. And by this process a textual summarization can be obtained. This textual summarization is encrypted by AES method and finally an encrypted summarization is obtained.

### B. Proposed system algorithm

Fig. 2, shows the flowchart of proposed system. Summarization is carried out by using LexRank Algorithm. And Encryption is carried out using AES Encryption method.

#### 1) LexRank Algorithm

Automatic summarization is the process of shortening a textual document, in order to create a summary with the major points or keyphrases of the original document. Generally, there are two approaches to automatic summarization: extractive and extractive.

Extractive summarization works by selecting a subset of existing words, key phrases, or sentences in the original text to form the summary. Whereas, abstractive methods build an internal semantic representation and then use natural language generation techniques to create a summary. Abstractive method creates the summary that is closer to what a human might express.

LexRank and TextRank are extractive summarization algorithms, which apply unsupervised graph-based ranking to build a summary. Essentially, they decide the importance of a sentence within a text. It is derived from Google's PageRank.

LexRank majorly consists of three steps:

- Preprocessing the text by removing the stop words.
- Creates a complete graph of which the vertices are the sentences of the text and the edges are weighted by the similarity between the sentences. The similarity is determined by the percentage of word overlap between sentences [7].
- Finally, the ranking function is executed on the graph and the highest scoring sentences are selected [7].
- LexRank uses cosine similarity of TF-IDF vectors as its similarity feature.

Let G = (V, E) be a directed graph. For a given vertex Vi, let adj(Vi) be the set of predecessors and successors of Vi. Given three vertices (sentences) R, S and T, the LexRank score (Erkan and Radev, 2004) is defined as:

$$IDF_{cos}(R,S) = \frac{\sum_{w \in R,S} f(w,R) \cdot f(w,S) \cdot IDF(w)^2}{\sqrt{\sum_{w_r \in R} f(w_r,R) \cdot IDF(w_r)^2} \cdot \sqrt{\sum_{w_s \in S} f(w_s,S) \cdot IDF(w_s)^2}}$$

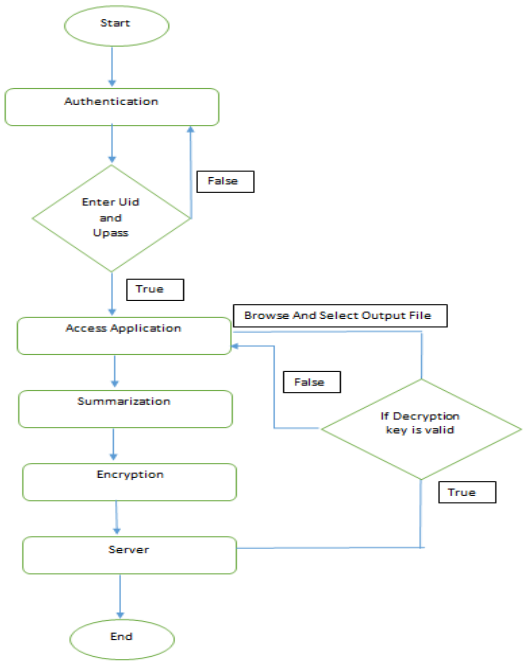


Fig. 2. Flowchart of proposed system

#### 2) AES Algorithm

The encryption process uses a set of specially derived keys called round keys. Following are the steps of AES encryption for a 128-bit block:

- Set of round keys is derived from the cipher key.
- State array is initialized with the block data (plaintext).
- Initial round key is added to the starting state array.
- Nine rounds of state manipulation are performed.
- Finally perform the tenth and final round of state manipulation.
- Copy the final state array out as the encrypted data or say cipher text.

Fig. 3 shows the flowchart of AES encryption method.

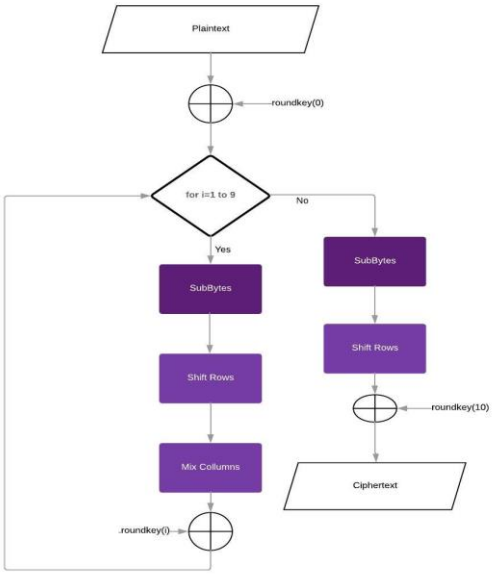


Fig. 3. AES Encryption method

## 4. Results and discussion

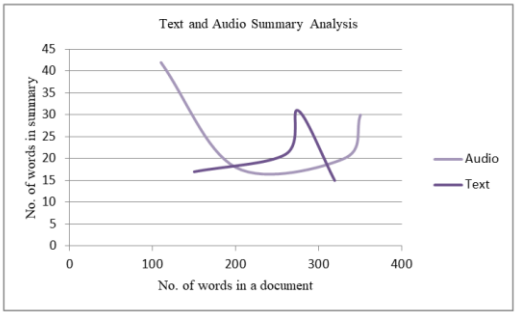


Fig. 4. Text and Audio Summary Analysis

Fig. 4 shows the summary analysis of textual and audio data. For text and audio files, the system aims to create textual summary. Audio file is first converted into text using IBM Watson Speech-to-text API. And then the summary is generated of converted audio file. As it can be figured out that summary result for text files are better than audio files. This is due to the reason that sometimes ill-transcriptions are formed. At maximum a summary of 30-40 words is formed. For text files this is count is on average 15-20 words.

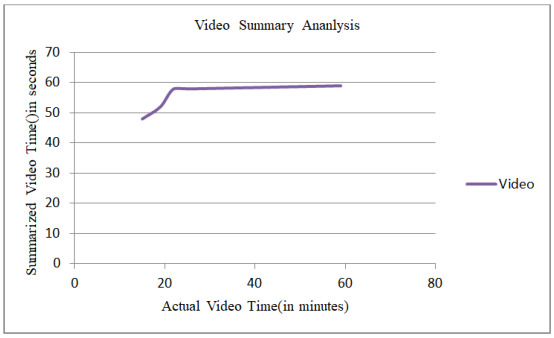


Fig. 5. Video Summary Analysis

Fig. 5 shows the summary analysis of video files. In contrast to text and audio files, for video files the system aims to create a small summarized video. From the graph it can be analyzed that video of about 48-59 seconds is formed, irrespective of the length of actual video. Further improvements can be done which takes into consideration the actual video length.

## 5. Conclusion

In this system, an asynchronous MMS task is being addressed, namely, a way to generate a textual summary from text, audio and video information. Readability is being addressed by selectively using the transcription of audio through guidance strategies. More specifically, a novel graph-based model is designed to effectively calculate the salience score for each text unit, leading to more readable and informative summaries [6]. Also the summary generated using this MMS is stored in server for further use. The summary stored in server is in encrypted form for the security issues. AES encryption method is used for encrypting data.

## References

[1] Erkan, Günes, and Dragomir R. Radev. “LexRank: Graph-based lexical centrality as salience in text summarization”. Journal of Artificial Intelligence Research (2004): 457-479.
[2] P. Li, J. Ma, and S. Gao. “Learning to summarize web image and text mutually”.Proceedings of the 2nd ACM International Conference on Multimedia Retrieval ACM, 2012.
[3] G. Evangelopoulos, A. Zlatintsi, A. Potamianos, P. Maragos, K. Rapantzikos, G. Skoumas and Y. Avrithis, “Multimodal saliency and fusion for movie summarization based on aural, visual and textual attention”.IEEE Transactions on Multimedia, vol. 15, no. 7, pp. 1553–1568, 2013.
[4] Fasila K. A. “Automated DNA Encryption Algorithm based on UNICODE and Colors”. IEEE International Conference on Electrical, Computer and Communication Technologies, Coimbatore, February 2016.
[5] Amal Joshy, Amitha Baby K. X., Padma S. and Fasila K. A. “Text to Image Encryption Technique using RGB Substitution and AES”. International Conference on Inventive Computing and Informatics (ICICI) 2017.
[6] Haoran Li, Junnan Zhu, Cong Ma, Jiajun Zhang and Chengqing Zong. “Read Watch, Listen and Summarize: Multi-modal Summarization for Asynchronous Text”, Image, Audio and Video”. IEEE Transactions on Knowledge and Data Engineering, June 2018.
[7] F. A. Grootjen, G. E. Kachergis, “Automatic Text Summarization as a Text Extraction Strategy for Effective Automated Highlighting,” 2018.