ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152

www.ijmece .com

Vol 12, Issue.2, 2024

A DEEP LEARNING BASED APROACH FOR INAPPROPRIATE CONTENT AND CLASSIFICATION OF YOUTUBE VIDEOS

[1]Ishrath Nousheen, [2]Najmunnisa Begum, [3]Momina Fathima,[4] Mohammad Nasera Thabassum

[1]Assistant professor in Department of Information Technology Bhoj Reddy Engineering College for Women

[2,3,4] UG Scholars in Department of Information Technology Bhoj Reddy Engineering College for Women

Abstract

Video information is one of the most emerging and easy ways to learn and know about anything that is going on around the world. On the internet, video material has grown in popularity influencing many parts of our life including education, entertainment, and communication. Video content is one of the most attractive ways where each and every individual attracts to the pictographic and visualizes the way content which helps in easy understanding and gain of knowledge. YouTube is the prime source for generating and classifying the text in which we have proposed the project implementation. YouTube is considered as of the most entertaining media where worldwide information is present. The main objective of our project is to generate and classify video content into different categories. We consider videos from YouTube which has subtitles. The primary goal is to extract and categorize information from videos. The procedure entails using NLP to extract text that may contain unwanted characters or symbols, necessitating text cleaning. The NLP is basically for analyzing the relevant information. To extract important information from text, certain text processing techniques such as tokenization and stemming may be required. The YouTube URL is copied and uploaded to the front-end web page. After the URL is uploaded the entire process takes place. The text generation using subtitles with NLP process where removal of stop words, and generating keywords is done through the dataset which is connected. The CSV file contains the records where the data is pre-processed and keywords are generated. Once the pre-processing is done,the summary is generated the retrieved text summary is categorized based on keywords and synonyms. The entire process is sent into the LSTM model to train and test the model for accurate output. Users can provide URLs, and the system will create a summary that is categorized appropriately. To give an interactive web-based output, the project is incorporated into the Flask framework.

I INTRODUCTION

Video metadata generation and classification and classification is the process of extracting relevant information from the input video content. Video metadata involves various forms in categorizing the information from videos such as their subtitles, tags,and duration as well as details about the content keywords, and topics. This categorization involves various

technologies to build a project. Nowadays videos and content from the video are luring people's attention in all aspects. From a 2 mins short to a full-length video everyone watches with utmost interest. The video content can be entertainment, news, sports, science and technology, education, etc. All the videos could not classify them into which category they belong our project gives a detailed explanation by generating a summary from which the classification takes place. For generating the text from the video we require Natural Language Processing (NLP). NLP is for analyzing text and generating relevant information. NLP is the subfield of Artificial Intelligence which focuses on the interaction between computer language and human language. Through NLP, computers can comprehend, interpret, and generate human language, especially with large volumes of data such as text, speech, and images. The main aim of NLP is to allow computers to process and analyze natural language data in a manner similar to humans, utilizing diverse techniques and algorithms, such as machine learning, deep learning, and linguistic analysis. The NLP involves keyword generation, removal of stop words, and extracting relevant information. The process involves summarization where YouTube API plays an important role in extracting subtitles from the video. The project we proposed is to extract the subtitles from which we generate the keywords. The text extraction is taking place through the NLP process. The major goal is to categorize data gathered from videos. To do this, we use NLP to extract text by obtaining the YouTube subtitle file using the transcript function. We next clean up the text by deleting any undesired symbols or letters.To extract relevant information, we may need to analyze the text further by doing tasks like tokenization, stemming, or entity recognition, depending on the unique use case Tokenization is the fundamental step in NLP for breaking down the text into each individual token. Tokenization and lemmatization play a vital role in pre-processing the given dataset which contains unnecessary data. These techniques are applied in order to retrieve the relevant information to generate a keyword dataset.The dataset is ready for testing and training of the model. The testing of the model is done through the LSTM algorithm where we achieve accurate data. The LSTM is capable of predicting outcomes based on short-term and long-term dependencies in the data, achieved by selectively retaining or discarding information as needed. After being trained, it can classify new videos by first extracting features from each frame with a CNN and then inputting these features into the LSTM for classification. The output presentation is in the form of a web page. The UI/UX design is used for designing of we-page. The URL is selected from YouTube and uploaded in the search box where the summary is generated. Through summary and detection of the keywords, we are able to classify the content

into specific categories. When the summary is generated based on the keywords matching and information-specific content is and efficient use.

categorized into sports, science, technology, news, entertainment, automobile, world, and

politics. Further, this categorization makes it easy for identification of content, accessible.

II LITERATURE SURVEY

a review of the current research on video metadata generation and classification, which employs a range of machine learning and deep learning techniques. The research draws upon various data sources, including pre-existing datasets available on platforms like Kaggle, in order to provide a comprehensive analysis of the field.“Efficient Video Classification Using Fewer Frames”, This research paper explains the fewer frames concept. Where the entire video is divided into each frame with a specific time slot. Where two datasets are compared by different algorithms and mathematical calculations to reduce the time for each frame. The paper introduces a unique approach that uses the notion of distillation to minimize the computing time necessary for video categorization. Training a teacher network, which constructs a video representation using all frames of the movie, is followed by training a student network, which analyses just a certain number of frames (k). Different loss function combinations are used to ensure that the student network's final representation and output probability distributions are similar to those of the teacher network. The suggested models are compared to a strong baseline and a skyline, with the results demonstrating that the proposed technique outperforms the baseline and offers a considerable decrease in computing time and cost when compared to the skyline. The method's success is illustrated on the YouTube-8M dataset, where the computationally less costly student network may cut calculation time by 30% while outperforming the instructor network.“Metadata extraction and classification of YouTube videos using sentiment analysis”, The research paper explains about the classification of video metadata can also be done by extracting and analyzing from a video followed by the decision-making of the content. Categorizing the video data into different titles which include the subject, URL of the video, Time limit of the video, video type, description, caption, etc. The information is segregated into different datasets to predict the accuracy of positive, negative, and neutral videos which thus calculates the rating of the video. The paper describes a method for evaluating YouTube video URLs that entails using sentiment analysis to determine the polarity of the video objects. The correctness of the entire process, however, is dependent on the Python inbuilt dictionary Corpus, which consists of a collection of terms and their accompanying scores.

The authors personally added new terms and their scores to the Corpus to evaluate its enhancement. As the dataset expands, automation will be required to make this process more efficient. Manually updating the Corpus dictionary would be time-consuming and inefficient. We suggest leveraging Machine Learning principles such as Neural Networks, Genetic Algorithms, SVMs, and Bayesian Learning to automate the process.

"Large-scale Video Classification with Convolutional Neural Networks", The current study looks at the effectiveness of convolutional neural networks (CNNs) in large-scale video classification. According to the findings, CNN architectures may successfully learn potent features from poorly labeled data, outperforming feature-based approaches in terms of performance. Deep convolutional neural networks (CNNs) were first used for video categorization in this landmark article. CNNs may be trained to recognize complicated visual patterns in films, such as object motion and scene changes, according to the authors. They also demonstrate that their model outperforms hand-crafted feature extraction approaches. Surprisingly the intricacies of the architecture's connectedness in time have no influence on this enhancement. A qualitative study of network outputs and confusion matrices reveals identifiable faults. Furthermore, the study demonstrates that a Slow Fusion model consistently outperforms early and late fusion alternatives. Furthermore, the study shows that a single-frame model can already achieve high performance, implying that local motion cues may not be necessary, even for dynamic datasets like Sports. Transfer learning tests on UCF-101 show that the taught traits are general and generic features, investigate approaches that consider camera motion, and explore the use of recurrent neural networks to combine clip-level predictions into global video-level predictions.

III EXISTING SYSTEM

There are various research papers with different approaches and implementations.

➢ To know whether the video is a positive, negative, or neutral video content sentiment analysis comes into the project implementation, where each YouTube URL is uploaded and then classified. A sentiment analysis technique is utilized in this system to identify the polarity of video objects. The performance of the system has been evaluated on a limited dataset.

➢ One of the interesting model buildings includes fewer frame concepts, where each frame of the YouTube video content is divided into a time slot. The fewer concept implementation if for a smaller memory footprint. The techniques used for classification include recurrent models, cluster-and-aggregate

models, and, memory-efficient cluster-and-aggregate models.

➢ Metadata to detect video action detection which includes multidimensional metadata for detecting human actions in the films. Video capturing, feature extraction, and acting performances are the key points in building a multidimensional model. To justify the building we use accuracy and recall techniques to fulfill the requirements.

➢ The objective of this study was to classify online videos based on their metadata. Web video metadata was extracted and stored in a database for the purpose of classification. The Random Tree and J48 classification algorithms were employed to classify the videos. The study found that the Random Tree classification model was more effective than the J48 model in categorizing online videos based on metadata.

➢ Video captioning is the task of generating natural language sentences that describe the contents of a video. Deep learning-based techniques are commonly used for this task, and there has been a significant amount of research in this area. According to the report, ResNet and VGG are widely used for extracting visual features, while 3D convolutional neural networks are extensively used for extracting spatiotemporal features. The project focuses on image and video captioning using deep-learning techniques.

➢ Video Search Engine: This system enables users to search for videos based on their metadata, which includes titles, descriptions, categories, and tags. It can also provide related video recommendations based on the user's viewing history and preferences.

➢ Video Recommendation System: This system uses machine learning algorithms to suggest videos to users based on their viewing habits and preferences. It can also generate metadata for new videos by analyzing user feedback and viewing patterns.

➢ Video Analytics Dashboard: This system provides insights into video performance and engagement metrics such as views, likes, comments, and shares. It can also generate metadata and summaries for videos based on their performance and audience engagement.

IV PROBLEM STATEMENT

The main aim of video metadata generation and classification is to produce a well-structured dataset that contains comprehensive information about the content of a video.This dataset can be utilized to organize, search, and retrieve videos efficiently.Toaccomplish this, the video content is analyzed, and descriptive details like the title, description, genre, language, and duration are generated. These details are then organized and classified to make searching and retrieving videos more convenient for users, based on

ISSN2321-2152
www.ijmece .com
Vol 12, Issue.2, 2024

specific criteria such as genre, language, or duration. Ultimately, the goal is to enhance the user experience by simplifying the process of finding the desired videos. Moreover, video metadata generation and classification can help content providers to gain insights into their audience and optimize their video content strategy based on user preferences and behavior.

V PROPOSED SYSTEM

The compilation and categorization of video metadata include collecting useful information from video footage and categorizing it based on specified criteria such as title, description, creator, creation date, duration, and tags. This is often accomplished by analyzing and extracting data from video using automated techniques such as machine learning algorithms and computer vision technologies. After that, the retrieved metadata's saved in a structured format for simple access and classification. Video classifications the process of categorizing video footage based on criteria such as genre, subject, audience, or language. It can be done manually or automatically using machine learning algorithms that analyze the video content and extract pertinent characteristics for classification. Overall, these activities are critical for effective video management and user experience enhancement. The project we proposed is to extract the subtitles from which we generate the keywords. The text extraction is taking place through the NLP process. The major goal is to categorize data gathered from videos. To do this, we useNLP to extract text by obtaining the YouTube subtitle file using the transcript function. We next clean up the text by deleting any undesired symbols or letters. To extract relevant information, we may need to analyze the text further by doing tasks like tokenization, stemming, or entity recognition, depending on the unique use case. After the text has been extracted from the video subtitles, it is classified by categorizing the generated summary. The classification is performed based on keywords and synonyms derived from the summary. The entire project is integrated to present the interactive web-based output through the Flask framework. Where the URL is uploaded, the summary is generated and classified

VI IMPLEMENTATION

- ***Video data collection:*** Collection of the dataset which contains the categorical

- ***Data Extraction***: Pre-processing the dataset to remove stop words, lemmatization, and stemming techniques through subtitles.

- ***Data generation***: After data extraction, using NLP generates the summary for classification.

- ***Training the model***: The entire model is trained using LSTM algorithm by labeled data.

ISSN2321-2152

www.ijmece .com

Vol 12, Issue.2, 2024

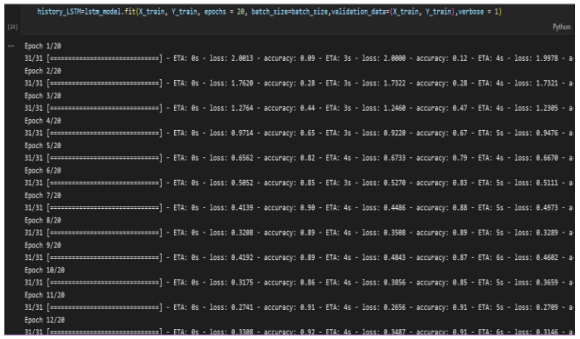
The training involves the optimization of labeled data.

➢ ***Testing the model***: The testing of the model involves the classification to predict the accurate output.

➢ ***Output:*** The entire model is presented in the form of a web-based application where the output screen is presented with the entire summary and classification. In order to build a successful model for video metadata generation and classification, it is crucial to collect relevant data. The dataset should be large enough and cover a wide range of categories and topics. It can be obtained from various sources, such as online video platforms or manually transcribed videos. To ensure the data is useful for the classification task, it should be categorized based on subtitles. Pre-processing the dataset is also important to remove noise and irrelevant information from the text. This involves removing stop words, punctuation, special characters, and numbers. The text should also be converted to lowercase for consistency. Techniques like lemmatization and stemming can be applied to reduce the text to its base or root form, making it more manageable and less complex.

To generate a useful summary from the pre-processed text, natural language processing techniques can be employed. The summary should be concise, easy to understand, and capture the most important information from the text. Extractive summarization selects the most important sentences, while abstractive summarization generates a new summary that conveys the essence of the text.

VII RESULTS

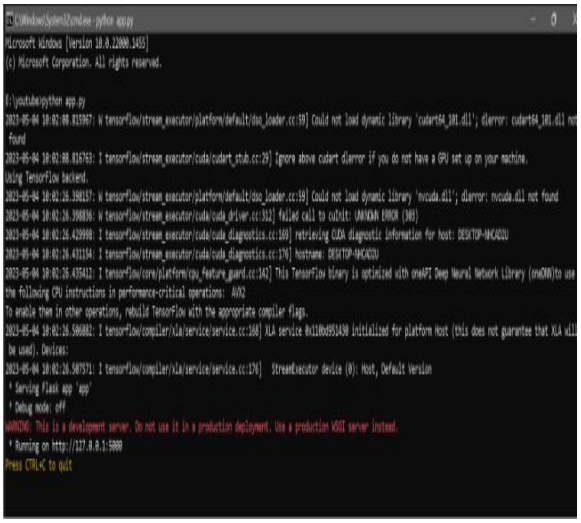


LSTM model and accuracy

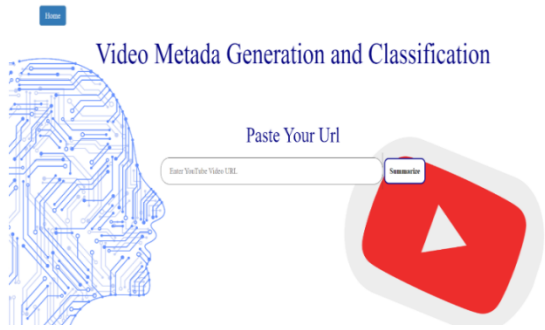
ISSN2321-2152

www.ijmece .com

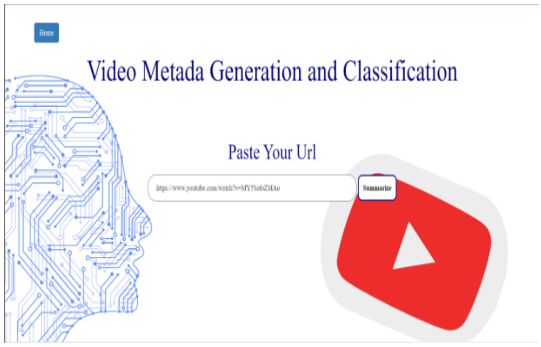
Vol 12, Issue.2, 2024



Running-Server



Front end application to upload URL



URL is uploaded for classification

Summary generation and classification

VIII CONCLUSION

The video metadata generation and classification using the NLP algorithm for text extraction and classification using the LSTM model is the prominent approach for accurate results. Natural language processing (NLP) and Long Short-Term Memory (LSTM) models are highly effective techniques for generating and classifying video metadata. NLP technique named Text summarization is a technique used to extract the most important information from a large document or a collection of documents. It can be used to create a summary of the document that captures the key points useful information can be extracted from video metadata such as world content, politics, news, sports, science, technology, and entertainment. This extracted information can then be used to train an LSTM model to classify videos into different categories By learning from a large dataset of labeled videos, the LSTM model can identify the patterns and features that are most significant for each category. This powerful combination of techniques can be used to enhance video search and recommendation

systems and gain a better understanding of user preferences and behavior. The function gets the YouTube API key from a configuration file. The function receives the video URL from the front end and extracts the video ID from the URL.The function makes a request to the YouTube API to retrieve the video details, including the video title and description. The function extracts the video transcript. If the transcript exceeds 2048 tokens, the function breaks and returns an error message. The function initializes the OpenAI API key. The function creates a prompt for the GPT-3 model. The function passes the prompt to the GPT-3 model and extracts a summary from the completion created by the model. The function removes any extra newline '\n' characters from the summary. Overall, this function leverages NLP and GPT-3 to generate a summary of a YouTube video based on its title, description, and transcript. This can be useful for quickly understanding the main points of a video without having to watch it in its entirety.

REFERENCES

1. Efficient Video Classification Using Fewer Frames", Shweta Bhardwaj, Mukundhan Srinivasan- NVIDIA Bangalore Mitesh M. Khapra

2. Metadata extraction and classification of YouTube videos using sentiment analysis, October 2016, Conference: 2016 International Carnahan Conference on Security Technology (ICCST).

3. Large-scale Video Classification with Convolutional Neural Networks, Andrej Karpathy; George Toderici; Sanketh Shetty; Thomas Leung; Rahul Sukthankar.

4. Automated Metadata Generation for Video Content" by X. Wang, L. Xie, and W.Zhang (2021)

5. Automatic Video Classification: A Survey of the Literature," in IEEE Transactions on Systems, Man, and Cybernetics, by D. Brezeale and D. J. Cook

6. A survey on video classification techniques by Nirav Bhatt published in 2015, Journal of Emerging Technologies and Innovative Research(JETIR)

7. Sentiment Analysis on YouTube: A Brief Survey, Published in the year 2015, by Muhammad Zubair Asghar, Shakeel Ahmad, Afsana Marwat, Fazal Masud Kundi

8. Video description: A comprehensive survey of deep learning approaches, Ghazala Rafiq, Muhammad Rafiq & Gyu Sang Choi, Artificial Intelligence Review (2023)

9. Metadata Based Classification and Analysis of Large Scale Web Videos, published in June 2015,International Journal of Emerging Trends & Technology in Computer Science

10. Video Classification:A Literature Survey, Published: Mar 31, 2018Pravina Baraiya,Asst. Prof. Disha Sanghani.