CATEGORIZATION OF YOUTUBE VIDEOS BY USING LSTM TECHNIQUE

Ms. S.V. LAKSHMI KUMARI, Assistant Professor, Department of E.C.E, Narayana Engineering College ,Gudur, SPSR Nellore, AP
.**G.Harshavi, M.V. Sai Charan, M.Bharath Kumar, .K.Sreenivas** Department of E.C.E, Narayana Engineering College ,Gudur harshavigottipalli@gmail.com

ABSTRACT : In this work, we will use text processing to classify YouTube videos as educational/non-educational. Many learners will benefit from the diverse educational content on YouTube. Watching instructional or educational videos only without interference is a challenge. An interesting video will catch our attention. Separating videos into educational and non-educational categories helps students focus on learning without being distracted. This paper introduces a mechanism for dividing videos into educational videos and non-educational videos. Classification is done by processing text from available keywords. In YouTube video metadata. These keywords are used to search YouTube videos. Our work innovated the Long Short-Term Memory (LSTM) method of video classification, which is a natural language processing technology that extracts keywords from metadata. This method can create an undisturbed learning environment for students faster and more effectively. The proposed method successfully divides videos into educational and non-educational categories.

Keywords: Educational Videos, Semantic Keyword Analysis, YouTube Content Analysis, Long short-term memory (LSTM).

1. INTRODUCTION

YouTube has over one billion active users and everyday hundreds of millions of people watch videos on YouTube and generate billions of views. It hosts over a billion videos with huge variety of content. YouTube is not only a hub for entertainment, it also has wealth of resources from lot of experts who share their knowledge for free. YouTube offers an opportunity to the students to learn new skills being taught by professionals and experts across the world. The video medium helps students in understanding and retention of knowledge. It provides the flexibility in learning at one's own pace. But the students can be easily provoked by the entertaining content in YouTube leading to distraction. Hence it is essential to develop a highly focused and efficient platform where users can browse through the educational content without being distracted. Most of the approaches used video processing in classifying the videos.

The research hypothesis of this paper is to consider the metadata from the given URL link to classify the videos and to develop a fast and accurate classifier. The video is processed and the keywords are extracted by the HTML extract text and processed by the text processing model. The cumulative result of this model is used in classifying the videos. This way the videos are classified as educational and non-educational. By doing this, a distraction-free environment can be created for the educational content on YouTube. This approach can be extended to all the categories of YouTube videos.

1.2. Text Processing

Text processing is the process of analyzing and processing text information, including extracting small pieces of information from text (also known as text extraction), assigning values or tags based on its content (also known as text sorting), or performing calculations that depend on text information. Because we naturally communicate in words rather than numbers, the company receives a large amount of raw text data through e-mail, chat rooms, social media and other channels. This unstructured data is full of ideas and opinions on various topics, products, and services. But the primary requirement of an enterprise is to organize, and measure text data in order to access this valuable information. One method of processing text data is manual, which is by far the most popular method. Natural Language Processing (NLP), a branch of AI, helps computers understand human language and extract value from text data.

Since text processing is one of the machine learning applications that ordinary technical consumers don't even know, most people use applications that have a text editor in the background every day.

2. LITERATURE SURVEY

There has been a lot of research on image and video classification . In our quick evaluate we most effective recognition on research which taken into consideration textual content functions.

Lin and Hauptmann describe an technique wherein units of functions – video and textual content – are used and SVM classifiers are skilled on the ones to expect whether or not a information video is a climate forecast or not. The paper introduces a unique manner of mixing predictions with the aid of using education a meta-classifier.

Feng et al. are possibly the first to apply co-education for photo annotation. They display that a smaller categorized set is needed to attain accuracy similar with that of a supervised learner. They additionally cope with the hassle of extracting applicable textual content functions from HTML pages. The classification assignment they do not forget includes assigning one in all 15 non-summary labels, including tiger, lion or cat to pics.

Cai et al do not forget the assignment of clustering photo seek results. The 3 types of functions they use are extracted from textual content surrounding pics, net web page hyperlinks and pics themselves. First, the textual content and hyperlink functions are used to become aware of companies of semantically comparable pics. After that, pics inside clusters are reorganized and clusters of visually comparable pics are formed.

Zhang et al classify movies with appreciate to five categories (movies, music, fun, finance and information) with the aid of using the use of binary classifiers skilled on separated characteristic units – meta-data (i.e., textual content) and content (i.e., visible). Their experiments verified that class accuracy relies upon at the sort of the classifier and consequently they take advantage of this prior knowledge by using a voting based class-established scheme. According to this scheme, the effectiveness of a classifier for predicting a positive class is expected at some point of the education segment and is later used at some point of classification.

Yang et al. gift a have a look at of video classification with respect to eleven catergories. They use 4 unique characteristic sources: visible, audio, textual content and semantic. The latter may be expressed with video annotations or "visible words". The textual content functions, which consist of video title, description and tags supplied with the aid of using the uploader, have tf.idf values. To amend the characteristic vector sparsity hassle, Yang et al. use WordNet-primarily based totally similarity degree to propagate the tf.idf amongst comparable words. More than 10K human-categorized movies have been used with inside the experiments. A binary classifier is skilled for each category feature combination. For predictions, ratings from all of the classifiers skilled to apprehend a positive class are fused to attain a final confidence.

3. EXISTING METHOD

If the video is located on the local hard drive/hard drive, perform the video classification task here. Convert the video to frames, and then apply image processing technology to each video. In the case of frames, the SVM classifier is used to classify the video. The toolbox provides supervised and unsupervised machine learning algorithms, including support vector machines (SVM), packed and guided decision trees, k-nearest neighbors, k-means, k-medoids, hierarchical clustering, Gaussian mixture models, and implicit Markov model. Machine learning statistics and algorithms can be used to calculate the amount of data that is too large to be stored in memory. Support Vector Machine (SVM) is a guided machine learning model that uses classification algorithms to solve two sets of classification problems. You can classify new examples by providing an SVM model of the labeled training data set for any category

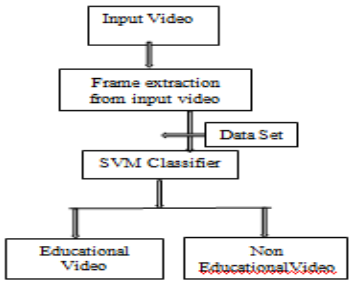

Fig 3.0 Block diagram of Existing Model

3.1. Support Vector Machine (SVM):

Support vector machine or SVM is one of the most popular supervised learning algorithms used for classification and regression problems; however, it is mainly used for machine learning classification problems. A better decision line or boundary can divide the n-dimensional space into classes so that we can easily put new data points into the correct classes in the future. This optimal solution limit is called the hyperlane SVM selects endpoints/vectors to help create hyperplanes. These edge cases are called support vectors, so the algorithm is called a support vector machine.

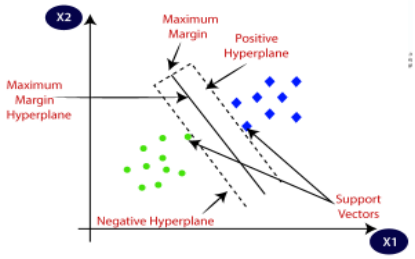

Fig3.1. Support Vector Machine

3.2. Working Of SVM:

The basic principle of support vector machines is to create a hyperplane that divides the data set into classes. Let's start with an example task. For example, suppose you need to separate the red triangle from the blue circle for a particular data set. our mission. is to create a line, divide the data into two categories, distinguish the red triangle and the blue circle. Although it can be assumed that a clear line separates the two classes, there can be many lines to accomplish this task. No one can agree that you are competent for this task. Let's visualize some lines that can distinguish these two categories, as shown below.

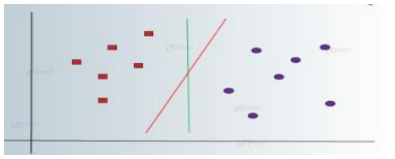

Fig 3.2. Working of SVM

According to SVM, we ought to locate the factors that lie closest to each the instructions. These factors are called guide vectors. In the following step, we discover the proximity among our dividing aircraft and the guide vectors. The distance among the factors and the dividing line is called margin. The purpose of an SVM set of rules is to maximise this very margin. When the margin reaches its maximum, the hyperplane will become the most efficient one. The SVM version attempts to expand the space among the 2 instructions via way of means of developing a well-described choice boundary. In the above case, our hyperplane divided the information. While our information became in 2 dimensions, the hyperplane became of one dimension. For better dimensions, say, an n-dimensional Euclidean Space, we've got an n-1 dimensional subset that divides the distance into disconnected components.

3.3. Disadvantages of Support Vector algorithm:

When classes within the data are points aren't well separated, which suggests overlapping classes are there, SVM doesn't perform well.

• we'd like to settle on an optimal kernel for SVM and this task is difficult.

• SVM on large data set comparatively takes longer to coach .

• SVM or Support vector machine is not a probabilistic model so we cannot explanation the classification in terms of probability.
• it's difficult to know and interpret the SVM model compared to Decision tree as SVM is more complex

4. PROPOSED METHOD

Here we are discussing about the Text Processing of the keywords available in the metadata of YouTube Videos using Long Short-Term Memory (LSTM), a Natural Language Processing (NLP) technique. Below figure shows the block diagram of our proposed model.

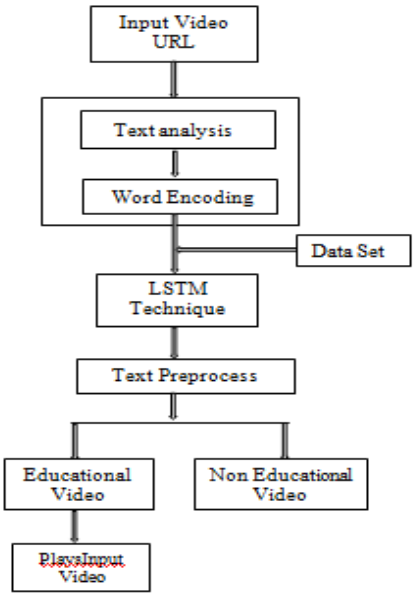

Fig 4.0 Block diagram of Proposed Model

4.1. LSTM Networks:

Long-term and short-term memory networks, often referred to as "LSTM", are a special type of RNN that can study long-term dependencies. They were introduced by Hochreiter & Schmidhuber (1997), and were improved and promoted by many people in subsequent work. They can solve a variety of problems well and are widely used today. LSTM is clearly designed to avoid the problem of long-term addiction. Memorizing information for a long time is largely their standard behavior, not something that is difficult for them to learn. All recurrent neural networks are a series of repeated neural network modules. In a standard RNN, this repeating module has a very simple structure, just like a single tanh layer.

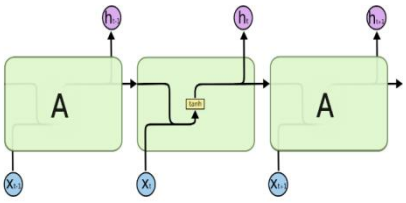

Fig 4.1. The repeating module in a standard RNN contains a single layer

LSTMs even have this chain like structure, but the repeating module features a different structure. Instead of having one neural network layer, there are four, interacting during a very special way.

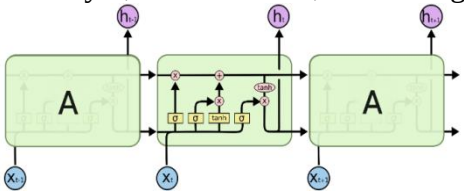

Fig 4.2 The repeating module in an LSTM contains four interacting layers.

In the above figure, each row carries a complete vector from the output of one node to the input of other nodes. The pink circles represent point operations, such as vector addition, and the yellow

rectangles represent the neural network layer to be checked. Connecting lines indicate connections, and forked lines indicate that your content is being copied and copies are being sent to different locations.

4.2. Introduction to NLP:

In this field of research, a series of useful applications is growing rapidly. They range from simple to complex. Some of them are:

- Spell check, keyword search, synonym search.
- Extract information from the website, such as: product price, date, location, person or company name.
- Classification: Reading level of school texts, positive/negative attitude towards longer documents.
- Automatic translation.
- Spoken dialogue system.
- Answer difficult questions.

In fact, these uses are widely used in the industry, from search (written and oral) to matching online advertising; from automatic/assisted translation to marketing or financial/trade sentiment analysis; from voice recognition to chatbots/dialogue agents (customers Service automation, equipment management, product ordering).

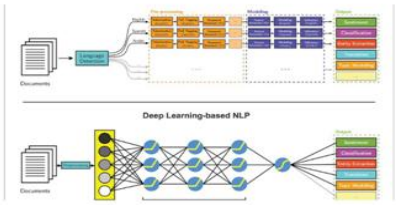

Fig 4.3. Deep Learning based NLP

4.3. Advantages of Proposed Method:

- Fast and good for data processing.
- LSTM is well-suited to classify, process and predict statistic given time lags of unknown duration.
- Relative insensitivity to gap length gives a plus to LSTM over alternative RNNs, hidden Markov models and other sequence learning methods.
- The constant error back propagation within memory cells lead to LSTM's ability to bridge very while lags just in case of problems.
- There appears to be no need for parameter fine tuning. LSTM works spill a broad range of parameters like learning rate, input gate bias and output gate bias.

4.4. The LSTM algorithm's update complexity per weight and time step is actually that of BPTT. this can be excellent as compared to other approaches like RTRL. LSTM is local in both space and time.

Applications of Proposed Method:

- Can be used in Educational system in categorizing the videos.
- Film Industries.
- Media Industry in differentiating the types of video.
- Police Departments in solving crime rates

5. RESULTS

The below figure window represent the input as a URL .

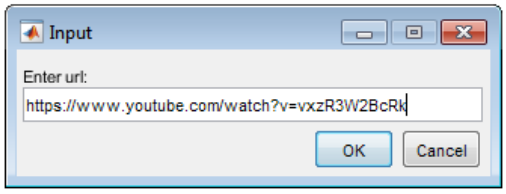

Fig 5.0. Input as a URL

The below histogram figure window represents the classification of data i..e., Educational and Non-Educational from the created data set

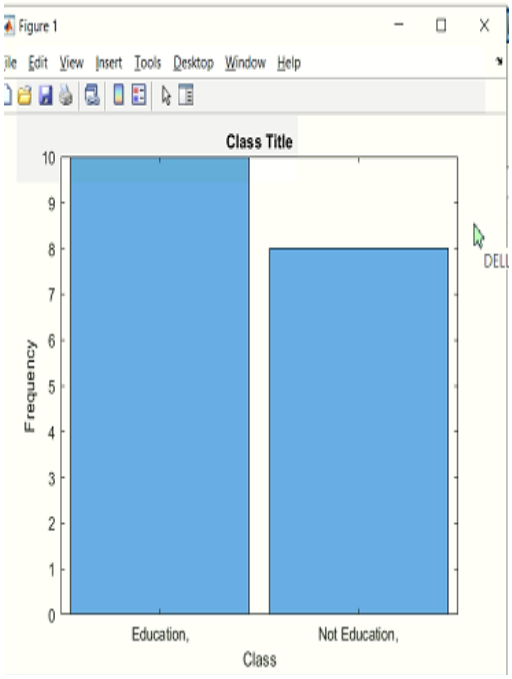

Fig 5. 2. Categorical Data

Training data is a very large data set used to train deep learning models**.** In the case of a monitored deep learning model, the training data is labeled. The data used to train the unattended deep learning model is not labeled.

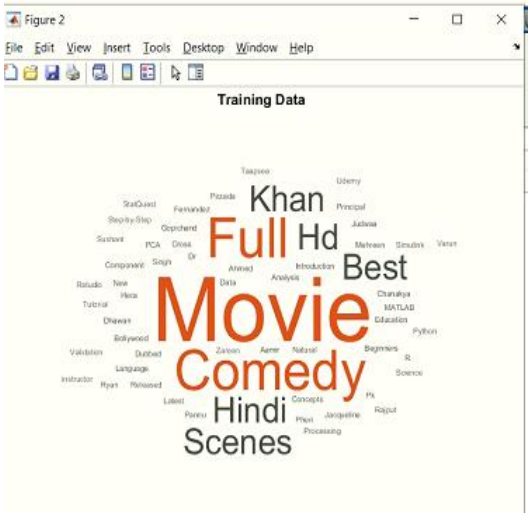

Fig 5.3.Training Data

To enter a document into the LSTM network, use word encoding to convert the document into a numerical index sequence. The training Options function provides options for automatically filling and truncating the input sequence. However, these options do not apply t**o** vector word strings. Instead, manually fill and trim the sequence. The below figure window represents the document length of the training data.

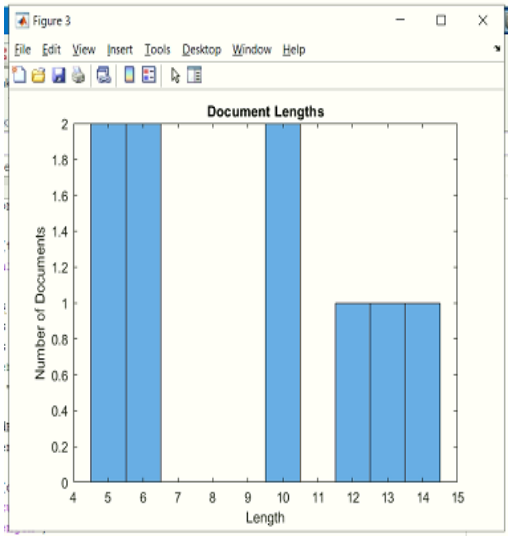

Fig 5.4. Document Length of Training Data

By creating various indicators during the training, you can understand the progress of the training. For example, you can determine whether the accuracy of the network has improved and how fast it has improved, and whether the network is consistent with the training data. The below figure the progress of training.

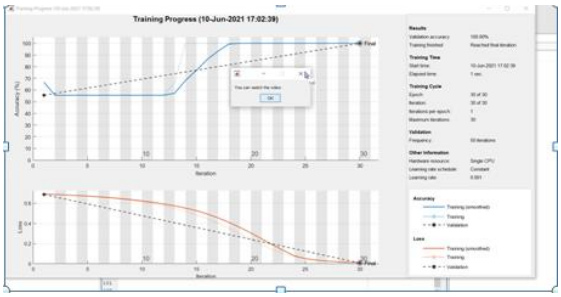

Fig 5.5. Training Progress

The Message box will be displayed as follows, when the given URL is an educational content and the video will be displayed as shown below.

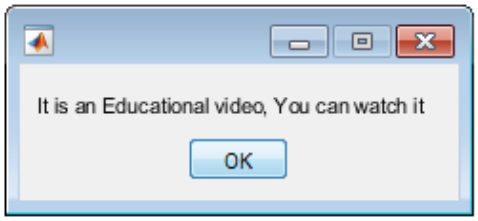

Fig 5.6. Output of Message box if video is Educational

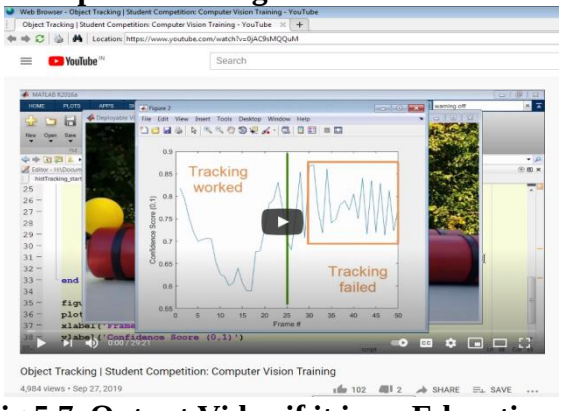

Fig 5.7. Output Video if it is an Educational

The Message box will be displayed as follows, when the given URL is an Non-educational content. We cannot watch the video if it is Non-Educational.

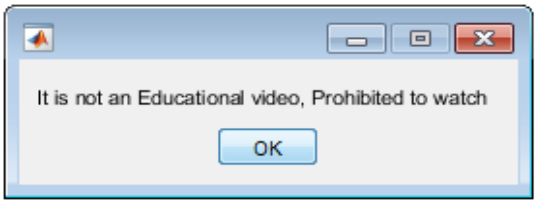

Fig 5.8. Output of Message box if video is Not Educational

6. CONCLUSION

In this project we presented a deep learning technique for Categorization of YouTube Videos as Educational or Non Educational. For this model, we have used LSTM (Long Short Term Memory) Network for the process of text processing. For the categorization of videos, some preprocessing Techniques are utilized for encoding the text from the given video URL. Experimental results shown that our proposed model provides better results when compared to existing works .

7. FUTURE SCOPE

In future, we can improve results of LSTM. And also we can extend this concept of categorizing videos in different fields including media, police department, and film industries with the help of deep learning techniques.

REFERENCES

[1] Aaron Chadha, Alhabib Abbas and Yiannis Andrepoulos:" Video Classification with CNNs: Using the Codec as a Spatio Temporal Activity Sensor." IEEE Transaction on Circuits and Systems for Video Technology Vol 29, No 2, February 2019.

[2] Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. "Large-scale video classification with convolutional neural networks". In IEEE Conference on Computer Vision and PatternRecognition (CVPR), pages 17251732, Columbus, Ohio, USA, 2014.

[3] S. Abu-El-Haija et al." YouTube-8M: A Large-Scale Video Classification Benchmark." aXiv:1609.08675be, 2016

[4]] S. Ioffe and C. Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". arXiv:1502.03167, 2015

[5] Video Classification with CNNs: Using the Codec as a Spatio Temporal Activity Sensor by Aaron Chadha, Alhabib Abbas and Yiannis Andrepoulos, IEEE Transaction on Circuits and Systems for Video Technology Vol 29, No 2, February 2019

[6] Classification of YouTube Data based on Sentiment Analysis by Shaila S.G, Prasanna MSM . IJRCSE, Vol 5, Issue 6, June 2018

[7] Large-scale video classification with convolutional neural networks" by A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei , In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, Ohio, USA, 2014.

[8] Improved Video Categorization from Text Metadata and User Comments by Katja Filippova, Keith B. SIGIR'11, July 24–28, 2011, Beijing, China, copyright 2011, ACM .