IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 5, May 2018

Video Sentiment Analysis using Multi Model Approach

Aishwarya Murarka[1], Kajal Shivarkar[2], Sneha[3], Vani Gupta[4], Prof. Lata Sankpal[5]

Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India[1,2,3,4]

Professor, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India[5]

Abstract: Sentiments are considered as the representation of human feelings and emotions. With the emergence of social media, people are now increasingly using the images, videos and audios in order to express their opinions on social media platforms. A growing source of consumer information is represented through Audio content that gained increasing interest from researchers, companies and consumers. Compared to traditional text content , audio-visual content provide a more real experience as they allow the viewer to better understand the reviewers emotions, beliefs and intentions through richer channels such as intonations. This article thus attempts to mine opinions and identify sentiments from the diverse modalities. A database of videos to be referred is used consisting a set of videos. The proposed system measures the speaker's current emotional state and requires at least 10 seconds of authentic speech to render the initial emotional analysis. All subsequent analyses are obtained every 5 seconds. The aim of multimodal data fusion is to increase the accuracy and reliability of estimates in turn to increase the usefulness of such systems in real-world applications.

Keywords: Sentiment analysis, Audio features, Text features, Multi-model classification.

I. INTRODUCTION

Nowadays, the age of Internet has changed the way people express their views, opinions. It is now mainly done through blog posts, online forums, product review websites, social media,etc. Social media is generating a large volume of sentiment rich data in the form of tweets, status updates, blog posts, comments, reviews, etc. Moreover, social media provides an opportunity for businesses by giving a platform to connect with their customers for advertising. Sentiment analysis (SA) tells user whether the information about the product is satisfactory or not before they buy it. Marketers and firms use this analysis data to understand about their products or services in such a way that it can be offered as per the user's requirements. Sentiment Analysis is the study of people's emotion or attitude towards event, conversation on topics or in general. Sentiment Analysis, shortly referred as SA, which identifies the sentiment expressed in a text then analyses it to find whether document expresses positive or negative sentiment. Majority of work on sentiment analysis has focused on methods such as Naive Bayesian, decision tree, support vector machine, maximum entropy. The approach followed in the paper investigates the challenges' and methods to perform audio sentiment analysis on video recordings using speech recognition. We use beyond verbalapi to transcribe recordings and describe emotion of speaker to identify the speakers involved in a conversation. Further, sentiment analysis is performed on the speaker specific speech data which enables the machine to understand what the humans were talking about and how they feel.

II. LITERATURE SURVEY

Normal-to-shouted speech spectral mapping for speaker recognition under vocal effort mismatch
This paper introduces a novel spectral mapping method which, when employed jointly with a statistical mapping technique, converts the Mel-frequency band energies of normal speech towards their counterparts in shouted speech. The aim is to obtain more robust performance in speaker recognition by tackling vocal effort mismatch between enrollment and test utterances.

A Study of Support Vector Machines for Emotional Speech Recognition
In this paper, efficiency comparison of Support Vector Machines (SVM) and Binary Support Vector Machines (BSVM) techniques in utterance-based emotion recognition is studied. Acoustic features including energy, Mel-frequency campestral coefficients (MFCC), Perceptual linear predictive (PLP), Filter bank (FBANK), pitch, their first and second derivatives are used as frame-based features.

Learning utterance-level representations for speech emotion and age/gender recognition.
Accurately recognizing speaker emotion and age/gender from speech can provide better user experience for many spoken dialogue systems. In this study, we propose to use deep neural networks (DNNs) to encode each utterance into a

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 5, May 2018

fixed-length vector by pooling the activations of the last hidden layer over time. The feature encoding process is designed to be jointly trained with the utterance-level classifier for better classification.

III. PROPOSED SYSTEM

The proposed system is a multimodal sentimental analysis system. Here we aim to analyze the parliamentary speeches though the system is flexible. A video is selected/browsed from the database . The video is then uploaded for analysis. The database comprises of at least 20 videos for the user to analyze.

The video is analyzed in two ways:
-Textual analysis
-Audio analysis.
Then the sentiments carried in the video by the speaker are displayed. Following is architecture diagram.

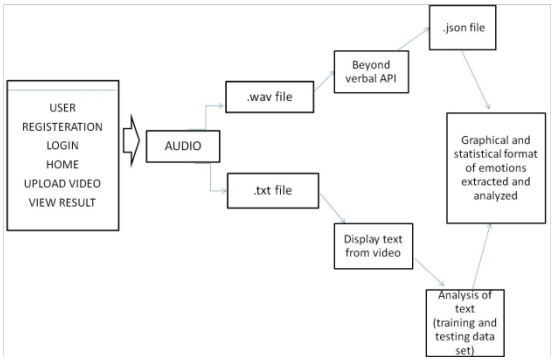

FIG : Block diagram

1)Audio Analysis-
-A .wav file is generated from the audio-visual clip(User uploaded video).
-The converted file is then transformed to meet the compatibility parameters of the API intended to be used.
-Video (.wav) is to be of the measures 8KHz, 16 bit mono as per the compatibility parameters for the input analysis.
- Compatible file is given as the input to the intended API(beyond verbal).
- Further a .json file is obtained carrying the raw analysis of .wav file.
- Contents of .json file associated with sentimental analysis:

- success/error
- Duration which states the time period of voice data processed in milliseconds
- SessionStatus which depicts started/processing/done.
- Analysis

-The 'Analysis' bit of json file consists of objects such as Temper, Valence and Arousal of the analysed wav file. These are the parameters generated after analysis of vocal biomarkers such as pitch, intensity and MFCC.
-The json file gives a more detailed analysis of these parameters by generating further objects

- Value -gives the magnitude of that parameter(Temper/Valence/ Arousal)
- Group- indicates the relevant group(low/medium/high/neutral) for duration of few sec.
- score - indicates confidence score (55-100) which is a metric that reflects the distributions of likelihoods of the identified output group (e.g. low arousal).

- These parameters(Temper, Valence, Arousal) are generated after every 1000 offset value.These parameters are further parsed by setting the threshold for their low, medium and high values and the values with higher confidence score are only considered.

- Values of low, medium and high , each represents different set of emotions. Further aggregate analysis gives the mean of valence, temper and arousal as well as frequent value of each group. At the end considering the frequent value of each group the mood can be classified.

a) Temper- Reflects the speaker's emotional state. It includes three main categories:

- Depressive
- Embracive
- Aggressive.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 5, May 2018

Its value ranges from 1 to 100.
-High Temper: Expresses aggressive emotions.
-Medium Temper: Embracive "positive" emotions and Self-controlled "neutral" emotions.
-Low Temper: Expresses negative emotions in an inhibited fashion.

b)Valence- Refers to the level of negativity or positivity.It ranges from 1 to 100.
-Negative Valence: The speaker's voice conveys emotional pain and weakness or aggressive and antagonistic emotions.
-Neutral Valence: The speaker's voice conveys no preference and comes across as self-control or neutral.
-Positive Valence: The speaker's voice conveys positive emotions, such as happiness, warmth, enthusiasm or calmness.

c)Arousal-Arousal is an output that measures the speaker's level of energy. It corresponds to similar concepts such as involvement and stimulation.It also ranges from 1 to 100.
-Low Arousal-The speaker's voice conveys less alertness and can be considered in cases of sadness, comfort, relief, or sleepiness.

-Neutral Arousal-The speaker's voice conveys a moderate degree of alertness and can be considered in cases of normal conduct, indifference, or self-control.
-High Arousal-The speaker's voice conveys a high degree of alertness such as excitement, surprise, passionate communication, extreme happiness, or anger.

2)Textual Analysis-
- The generated wav file is further converted to text.
- Along with the text, beyond verbal generates phrases which are the adjectives and adverbs representing the video, are given as input for textual analysis.
-These two inputs are analyzed using keyword extractor function which identifies most important word within a document.
- It takes inString List required text to analyze Integer which only return some of the most likely topics, threshold Float optional only return topics with likelihood greater than a specific number, relative which is Boolean (defaults to False) -when True, it will scale the scores for each keyword such that the lowest is 0 and the highest is 1, which makes setting a threshold easier.
- This function will return a dictionary with top_n key-value pairs. These key-value pairs represent the likelihood that each of the extracted keywords are relevant to the analyzed text.
- The keys in the dictionary are strings containing the extracted keywords, and the values are the likelihoods that these keywords are relevant to the analyzed text.

IV. EXPERIMENTAL SETUP

- User interact with the system using web application where user need to face the Login page to enter the login credentials.
- If the user does not have the login credential or for the first time user need to register to the proposed system by entering basic details like Name, address, contact number, email-id and password for login.
- After login they will get the respective options for interacting with the system and user will get options to perform the basic functionality according to the requirement.
- Application will use Java language for development of the project. My-SQL is used as a backend language.
- Technology :

- Java
- J2EE
- Apache Tomcat server
- Java Version is J2SDK 1.7 / 1.8

- Database used is My SQL

V. RESULT

.Result for the textual analysis:The textual analysis output can be represented in the following categories:
-Positive Sentiments: Happy, joy,surprise.
-Negative Sentiments: anger, aggressive, sad.
User will register to application. After registration, user will get user id and password. Using login credentials user is able to login to application. It is shown in fig. 1(a) and (b).

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 5, May 2018

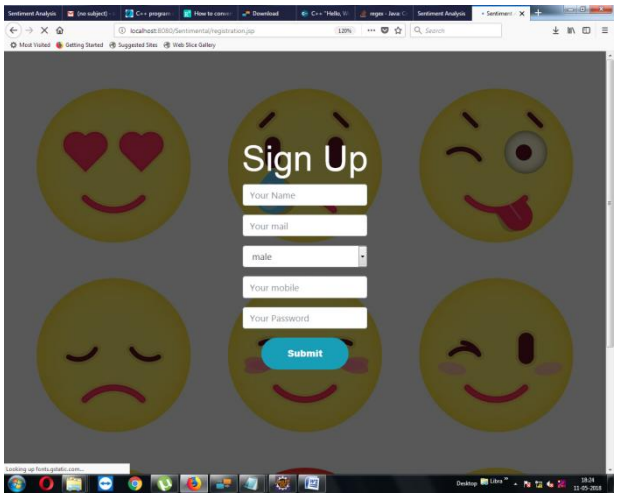

FIG.1(a)

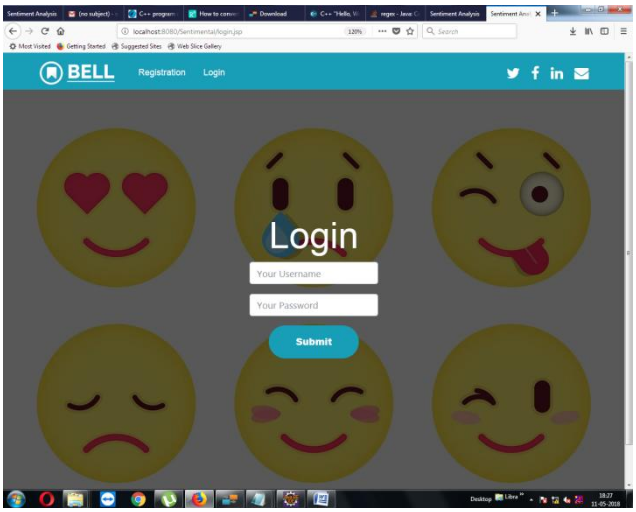

FIG. 1(b)

User will browse video. After browsing user can click on start analysis. Analysis of video is done on basis of textual and audio analysis. It is shown in fig. 2 and 3.

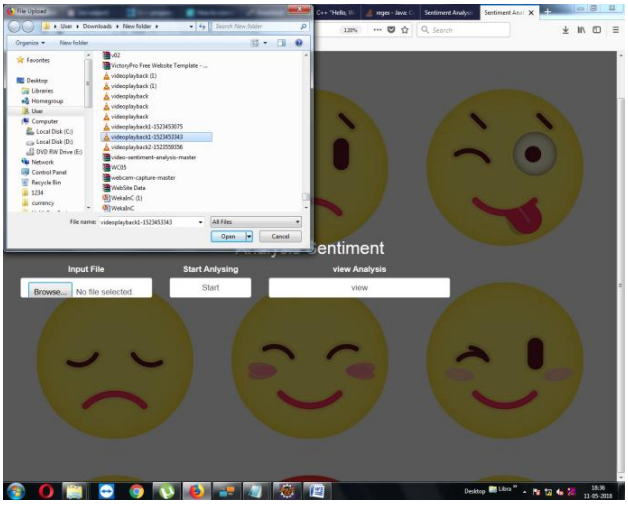

FIG 2

DOI 10.17148/IJARCCE.2018.7549

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 5, May 2018

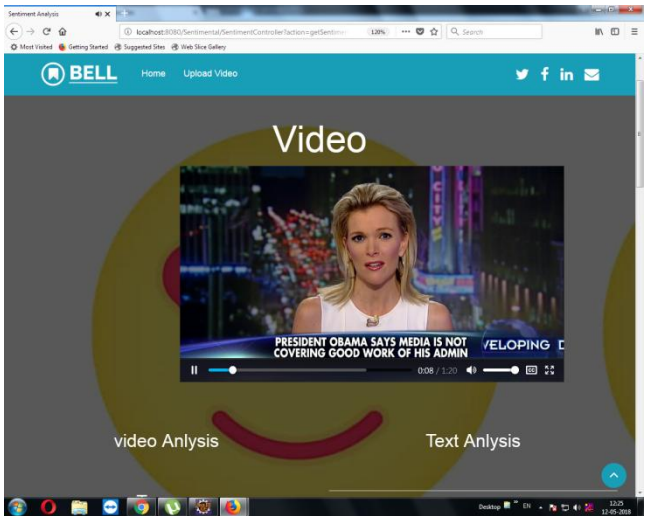

FIG. 3

Result for the audio analysis:

The audio analysis output can be represented in the following categories:

- Sadness/Uncertainty/Boredom – As a measure of low arousal.
- Dislike/Anger/Stress- As a measure of high arousal,
- Neutral-The state of these emotions in terms of arousal is neutral.

The analysis contains the confidence score of the following broad mood groups:

- Temper- consists of three distinct groups:
- Low
- Med
- High

Result is shown in fig. 4

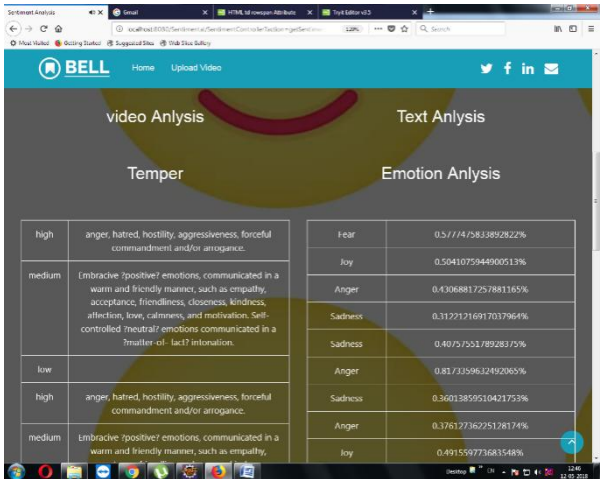

FIG.4

Graphical representation is shown below. We have analyzed video and text of given uploaded video. Arousal, temper, valence are parameter for representation. In text analysis, we have chosen fear, joy, anger. It can be shown in following figure.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 5, May 2018

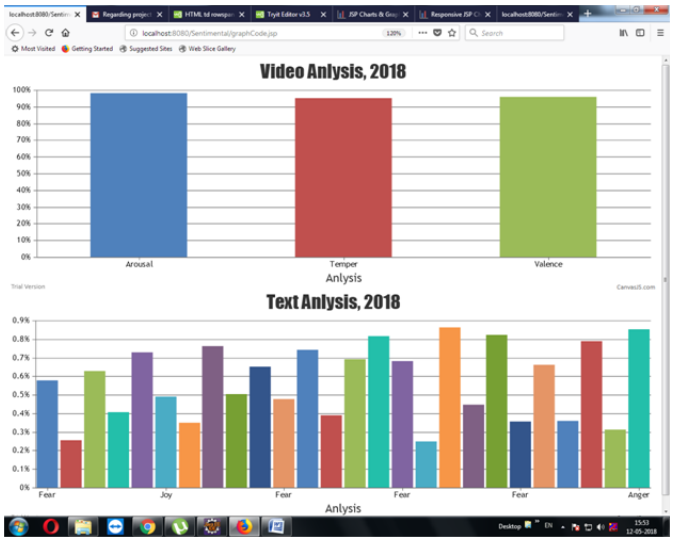

FIG.:- Graphical representation

VI. CONCLUSION

We conclude that sentiment analysis is a wide area of research and could be used in variety of applications. Our system tends to provide a reasonable amount of precision for the value of distinct sentiments. We believe that multimodality will also help in detecting whether a speaker is expressing his own opinion or merely parroting somebody else's views. In such cases a mere text based approach will fail, as the most important clues will be found in intonation and facial expressions. Multimodal SA is very much an open ended topic. Lots more research needs to be done as evident from the results of the discussed experiment.

REFERENCES

[1] Nattapong Kurpukdee , Sawit Kasuriya , Vataya Chunwijitra, Chai Wutiwiwatchai and Poonlap Lamsrichan ," A Study of Support Vector Machines for Emotional Speech Recognition", 978-1- 5090-4809- 0/17/$31.00 ©2017 IEEE
[2] Harika Abburi," Audio and Text based Multimodal Sentiment Analysis using Features Extracted from Selective Regions and Deep Neural Networks", International Institute of Information Technology Hyderabad - 500 032, INDIA June 2017
[3] Zaher Ibrahim Saleh Salah," Machine Learning and Sentiment Analysis Approaches for the Analysis of Parliamentary Debates", May 2014
[4] "Towards real time sppech emotion recognition using deep neural network " 2017
[5] Lakshmish Kaushik , Abhijeet Sangwan, John H. L. Hansen," SENTIMENT EXTRACTION FROM NATURAL AUDIO STREAMS", 978-1-4799-0356- 6/13/$31.00 ©2013 IEEE
[6] S. Lugović, I. Dunđer and M. Horvat,"Techniques and Applications of Emotion Recognition in Speech", MIPRO 2016, May 30 - June 3, 2016, Opatija, Croatia.