ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

Hybrid Approach For Emotional Speech Classification

C.Jayashree[1] ,V. Dharani[2], K. Abiram[3], R. Ragaventhirah[4] , K. Suvidha[5]

Department of CSE, Dr. Mahalingam college of Engineering and Technology, Pollachi, Tamil Nadu, India

ABSTRACT: In today's world most of the product reviews and advertising campaigns are made through videos and identifying emotions from those videos are quite unexplored field of research. The traditional approach for emotion classification involves obtaining audio from the videos and extracting the text from the audio using Speech Recognition system and the obtained transcript is used for emotional analysis. Emotional analysis of a speaker using traditional approach consumes larger span of time and involves difficulties in classification of the speech using only text or audio data. In the proposed system a weight age-based model is developed that produces the result combining the results of emotion recognized from audio input and emotion recognized from the textual input, outperforming the traditional systems.

KEYWORDS: Emotion, Audio, Video, Reviews.

I.INTRODUCTION

Emotional analysis is useful in extraction of people's view over certain products or items. This is very popular on internet campaign videos which could be useful in predicting people's opinion about the product. On a daily basis, millions of people express their views on products, services and offers, among others, using online platforms such as social networks, blogs, wikis, discussion boards, etc. Emotional analysis is valuable towards enhancing sales and improving a company's marketing strategies by tracking customer reviews and survey responses, identifying ideological shifts and analyzing trends in political strategy planning, financial reports and recorded social media sentiments.

People express emotions not only verbally but also by non-verbal means. Non-verbal means comprise body gestures, facial expressions, modifications of prosodic parameters, and changes in the spectral energy distribution. Often, people can evaluate human emotion with only the speaker's voice since intonations of a person's speech can reveal emotions. Extraction of emotions from the videos are developing field of research, most of the traditional techniques involves extraction of audio and classification of emotions from the transcript of the audio. C.Nagarajan *et al*.[2,7,11] The obtained text is mostly subjected to Natural Language Processing (NLP) processes (such as stemming, part-of- speech tagging, etc.) with utilization of additional resources (e.g. thesauri, sentiment or emotion-based lexicons, sophisticated dictionaries and ontologism) to model the documents towards a successful sentiment detection. Since the process becomes clumsy on extraction of audio and then performing emotional analysis on the obtained transcript, the proposed system extracts text and audio using the URL of a video and uses a weight age-based model to identify the sentiments determined from both text and audio.

Rest of the paper is organized as follows, Section II contain the related work of the existing literature on this topic. Emotion detection from speech, Sentiment analysis on speaker data and the Current state of text sentiment analysis are discussed. Section III explains the Emotion classification methodology with flow chart, Section IV describes results and discussion and Section V concludes research work with future directions.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

II.LITERATURE SURVEY

The sudden eruption activity in the area of opinion mining and sentiment analysis, which deals with sentiment, subjectivity and opinion in text. It has occurred in part as a direct response to the surge of interest in new system that deals directly with opinion as a first-class object.[1] Sentiment analysis is one of the fields in affecting computing refers to all the areas of detecting and evaluating human sates of mind towards at different situations, events or any other interest. Opinion from the texts can be identified by using Opinion-mining tasks. Opinion mining tasks are subjectivity detection, opinion polarity classification, opinion spam detection, opinion summarization and argument expression detection. Emotion from text can be derived by using Emotion-mining tasks. Emotion mining tasks are emotion detection, emotion polarity classification and emotion cause detection. The features used to find the usefulness of many feature and techniques from a large part of literature they are Presence-Based and Frequency-based features, Unigram and N-Gram Features, Part of Speech, Syntax, Negation the polarity of a sentence can be changed totally by the use of negating words, Topic-Oriented Features. Social Information Processing theory states that verbal clues are used instead of non verbal clues in computer-mediated communications. The non verbal clues are used in face-to-face environments. Building a lexicon is a tedious and time-consuming task, automatic solutions called "lexicon expansion "methods. Polarity of new words are used by propagating the seed word. Contextual polarity is the word with respect to the context. Label transferring is used in some of the prior works of domain adaptation. In every domain the cluster can be featured into two groups i.e,Domain independent and Domain specific features[2] Internet is the major resource place of sentiment information where people post their opinion through many social media. For research purpose the websites release many API's to collect the prompting data which is used by researchers. The different collected API's are used to create new applications. Hence online data plays major role in sentiment analysis. Categorization of sentiment polarity is a problem in sentiment analysis where a text can be categorized into specific sentiment polarity. The three levels of sentiment polarity categorization are sentence level, document level, the entity and aspects level.[3]The information from the textual document is extracted with the help of sentiment analysis and opinion mining. The information for analysis are gathered from online posted product and service reviews which used for business and marketing. There are number of tools, libraries ,API's are used for sentiment analysis . The tools like Stanford coreNLP the framework for performing NLP tasks. The platform used to build python program is Natural Language Toolkit that utilize human data. Based on the text classification the sentiment analysis tool classify the sentence into positive, neutral and negative. The Lexicon-based extraction feature extraction method is based on sentiment lexicon which consist of set of terms in specific language, carry emotional weight along with a number of dimensions. Term can be classified according to their subjectivity i.e. classification happens either subjective or objective or their polarity. The datasets are transformed into vectors whose size and forms are depends on the selected lexicon and vectorization type. For sentiment prediction the derived vectors are processed by selected classification algorithm. Vectorization schemes are Bag of words representation, Average Emotion representation, Mixed representation.[4] Speech transcript from speaker conversation is used to detect speaker's emotions. In sentiment analysis most of the work has focused on methods like Naïve Bayesian, Decision tree, Support vector machine, Maximum entropy. Speech recognition is the process of identifying words in the speech spoken by the speakers and convert those data into machine understanding format which can be used later by machine. The tool used to recognize speech is Sphinx.It is the unique characteristics. Speech as signal contains many features which contains emotions, linguistics, speakers specific information. For designing speaker discriminate system Mel Frequency Cepstrum Coefficient is used. For better accuracy the extraction of speaker discriminate feature is important. MFCC act as a human ear and acoustic frequencies are mapped to Mel frequencies. Dynamic Time Wrapping techniques is used for feature matching. This technique is used to measure the similarity with help of minimum distance between them.[5]Emotions are so complex however we can't put all emotions into an particular emotion category, most of the emotional states can be described as a mixture of multiple emotions. Various classification algorithm is used in emotion recognition in speech such as Hidden Markov Model and Neural Network. Kinds of basic emotions are classified into four types-neutral, anger, happiness and sadness. Using telepays we need to collect large amount of emotional speech samples. Vocal energy and speaking rate contribute to vocal emotion signaling. Pitch derivative and pitch slopes are the features of pitch. If the slope is upslope it is positive, or else it is negative.[6]Apart from confusion matrix,the transition matrix and emission matrix produces the sequence of states. Sentiment analysis application used to analyse the attitude

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

of individual debaters. According to emotional studies the tone of every voice can be characterized by its pitch, loudnes and speech rate. Particularly, automatic speech recognition (ASR) of natural audio streams and text spoke in audio is difficult and the resulting transcripts are not very accurate. The difficulty stems from a variety of factors including (i) noisy audio due to non-ideal recording environments, (ii) foreign accents, and (iii)spontaneous speech production. The sentiment extraction mainly on text and videos.On text datasets, This provides us with the capability of identifying key words/phrases within the video that carry important things. On indexing these key words/phrases, retrieval systems can enhance the ability of users to search for relevant information faster. Mainly the audio features uses like pitch, intensity and loudness are extracted using Open- EAR software and Support Vector Machine (SVM) classifier is built to detect the sentiment. There are two types of feature extraction they are frame-based and itterence based features. Features like energy, pitch and Mel frequency cepstral coefficients are extracted as frame based features, where utterance based features are calculated by statistical values like maximum, minimum and mean variance of those frame based features. All these extracted features are concatenated together to get a better result of accuracy.[7]The emotional speech data produced by actors is analyzed by using the segmental and temporal changes which are caused by humans articulary behaviour accompanied by emotion arousal. The degree of of segmental changes were detected by the extent of emotion reductions and this describes the deletion and assimilation of segements .
This produces the results like namely reduction effort, gestural recognition, cognitive constraints and speaking styles. The process which are done in assimilation and deletion technique is two label files were created using spectrographic, oscillographic and auditory analysis. The first label file will represent the segmental structure and narrow transcription of uttrences whereas the second label file represents the syllable structure of sentence and categorization of each syllable. The emotions can also be detected using the vowels in the text. Regarding to vowel formant analysis observed that sentences expressing fear sadness and boredom are characterised by formant shift toeards centralized position in different places of vowels

III. MODULE DESCRIPTION

Video Fetcher and Preprocessor:

The input is given as an URL from which videos can be extracted using python module youtube-dl which helps in extraction of audio and text. The extracted text is processed for removal of noise data including unicode symbols and other unrelated characters. Similarly the extracted audio file is processed for ffmpeg converter which processes the audio to desired format i.e WAV (Windows Audio Video File).

Emotion Classifier-Audio:

In this module the generated wav files are given as input to SVM Classifier which classifies the emotions based on the audio features including pitch, frequency and Mel Frequency Cepstral Coefficient.

Emotional Classifier -Text:

The extracted text content is subjected to removal of stop words and is subjected to Natural Language Processing using Python's NLTK for classification of emotions.

Emotion Classifier-Hybrid:

Hybrid approach uses the results of both audio and text and helps in proposing a suitable method for classification depending on the accuracy of individual methods.

IV. METHODOLOGY

The input URL is validated and processed through the preprocessor. The preprocessor then fetches the video and downloads the audio along with its transcript. The transcript is cleaned by removing the noisy data involving upper-case alphabetic characters converted to lower-case letters followed by numeric digit removal. The obtained content is then processed using the classifiers to obtain the result.

Figure 1 shows the overall block diagram of the process. The obtained text transcript and audio are fed to different classifiers. The Emotion Classifier - Text uses the obtained transcript and classifying the emotions based on lexicon

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

weight. The Emotion Classifier - Audio pre-processes the audio removing silences and uses SVM classifier for obtaining the emotions. Finally, the hybrid classifier generates a weightage model on results from both classifiers producing the final result. The implementation was done in Windows 10 environment using Python.

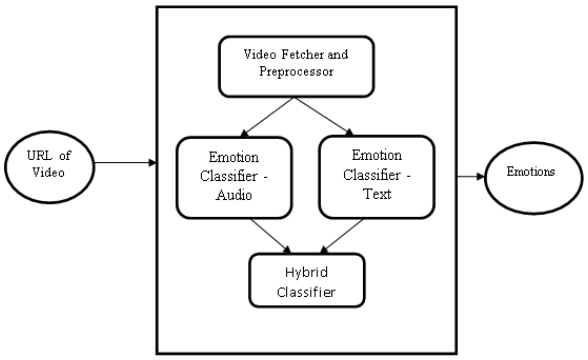

Figure 1 Block Diagram of Emotion Classifier

A. Pre-processing of input data

The input data is both text and audio. For text the stop words are removed and words are converted into lowercase. We have to divide speech signal into frames. Then we compare each frame with phonemes label in the database and find the frame have a silence phoneme label and remove that frame. After that, we merge whole speech frames into utterances again. Silence is considered as a useless data in this research. After getting speech data, all of them are divided into frames. These frames are processed using the classifier.

B. Emotion Classifier – Text

The classifier uses lexical based approach using the dictionary of sentiment. We use this dictionary to assess the sentiment of phrases and sentences, without the need of training a model. Sentiment is categorized as {negative, neutral, positive} and it is found be numerical like a range of intensities or scores. Lexical approaches look at the sentiment category or score of each word in the sentence and decide what the sentiment category or score of the whole sentence is. The result is obtained as three numerical values denoting the

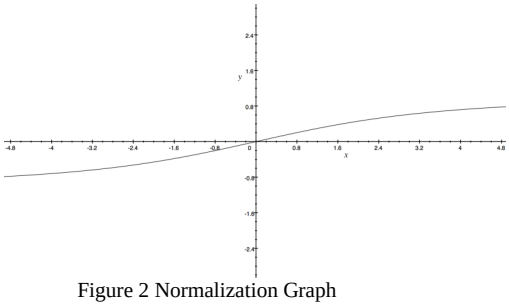

Figure 2 Normalization Graph

C. Emotion Classifier – Audio

Speech signal data is feed to the system as an input to the system. As the input sound contains noise signal/silent zone, signal pre-processing of the signal is required to chop the silent zone after pre-processing of the signal the spectral analysis is done. The next stage of the system is to extract the speech features like Formant Frequencies, Entropy, Median, Mel-Frequency Cepstral coefficient, Variance, Minima etc. from the filtered emotional speech signal .Some of the speech extracted features may be redundant or even cause negative effects to the training of neural network for that feature selection method is applied, through which only that features which adds efficiency to the system is chosen so as to build an efficient system with greater accuracy. After selection of feature vector, a feature database is built up this is required as an input to

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

classifier. On the basis of this database classifier which is vigorously train on the given input to recognize Human emotions with the accuracy.

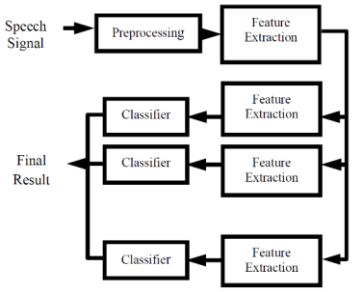

Figure 3 Speech Emotion Classifier

Figure 3 shows the visualization of the audio classifier where each feature is extracted separately to produce the final result.

V. RESULTS AND DISCUSSION

A. Evaluation Metric

The precision and recall and F-measure values for both the systems are formed.

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{recall} = \frac{TP}{TP+FN}$$

$$\text{F1} = \frac{2 \times precision \times recall}{precision + recall}$$

$$\text{accuracy} = \frac{TP + TN}{TP + FN + TN + FP}$$

	Positive	Negative	Neutral
Positive	TP	FN	PP
Negative	FP	TN	PN
Neutral	PP	PN	PNeutral

Used to check the accuracy. It is very useful in hybrid approach.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

B. Summary of Results

Thus, it has been identified that the weightage-based hybrid model is needed for the effective analysis of sentiments asthe text-based classifier produced only 81% accuracy.The presence of audio classifier improves the performance of the sentiment classifier by a large margin on the benchmark dataset. With a hybrid classifier, it is able to predict the sentiments with higher accuracy compared to the text-based analysis.

VI.CONCLUSION

Thus the proposed system a weight age-based model is developed, that produces the result combining the results of emotion recognized from audio input and emotion recognized from the textual input, outperforming the traditional systems.

REFERENCES

[1] Bo Pang and Lillian Lee, " Opinion mining and sentiment analysis", Yahoo! Research, 2011.
[2] E Geetha, C Nagarajan, “Induction Motor Fault Detection and Classification Using Current Signature Analysis Technique”, 2018 Conference on Emerging Devices and Smart Systems (ICEDSS), 2nd and 3rd March 2018, organized by mahendra Engineering College, Mallasamudram, PP. 48-52,2018
[3] Xing Fang* and Justin Zhan, " Sentiment analysis using product review data", Journal of Big Data, 2015.
[4] Maria Giatsoglou, Manolis G. Vozalis, Konstantinos Diamantaras, Athena Vakali, George Sarigiannidis, Konstantinos Ch. Chatzisavvas, " Sentiment analysis leveraging emotions and word embeddings", M. Giatsoglou et al. / Expert Systems With Applications, 2017.
[5] Maghilnan S, Rajesh Kumar M, " Sentiment Analysis on Speaker Specific
Speech Data", International Conference on Intelligent Computing and Control (I2C2), 2017.
[6] Feng Yu, Eric Chang, Ying-Qing Xu, Heung-Yeung Shum. Emotion Detection From Speech To Enrich Multimedia Content, Microsoft Research China, 2011.
[7] C. Nagarajan, M.Madheswaran and D.Ramasubramanian- ‘Development of DSP based Robust Control Method for General Resonant Converter Topologies using Transfer Function Model’- Acta Electrotechnica et Informatica Journal , Vol.13 (2), pp.18-31,April-June.2013
[8] Miriam Kienast, Walter F. Sendlmeier, " Acoustical Analysis Of Spectral And Temporal Changes In Emotional Speech ", Technical University Berlin, Institute of Communication Science, Germany, 2000.
[9] Dimitrios Ververidis and Constantine Kotropoulos, " Automatic Speech Classification To Five Emotional States Based On Gender Information", Artificial Intelligence and Information Analysis Laboratory, Greece, 2003.
[10]Lakshmish Kaushik, Abhijeet Sangwan, John H. L. Hansen, " Sentiment Extraction From Natural Audio Streams", Center for Robust Speech Systems (CRSS), Eric Jonsson School of Engineering, Texas, U.S.A. , 2011.
[11]C.Nagarajan and M.Madheswaran - ‘Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques’- Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011
[12]Muzaffar Khan, Tirupati Goskula, Mohmmed Nasiruddin, Ruhina Quazi, "Comparison between k-nn and svm method for speech emotion recognition", International Journal on Computer Science and Engineering, 2011.
[13]Emily Mower, Shrikanth Narayanan,"Emotion Recognition Using a Hierarchical Binary Decision Tree Approach", Signal Analysis and Interpretation Laboratory, 2009.
[14]Lingli Yu, Kaijun Zhou, Yishao Huang, "A Comparative Study on Support Vector Machines Classifiers for Emotional Speech Recognition, "Immune Computation , Number1, March 2014.
[15]Norhaslinda Kamaruddin, Abdul Wahab, "Heterogeneous Driver Behavior State Recognition Using Speech Signal ", Conference Paper, October 2011.