IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# AUDIO SENTIMENT ANALYSIS

[1] P.Ansar khan, [2] T.sumanth, [3] K.vishnu vardhan

[1] Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Anand Nagar, Krishnankoil, India

[2] Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Anand Nagar, Krishnankoil, India

[3] Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Anand Nagar, Krishnankoil, India

***Abstract:*** Audio sentiment analysis using automatic speech recognition is an emerging research area where opinion or sentiment exhibited by a speaker is detected from natural audio. It is relatively under-explored when compared to text based sentiment detection. Extracting speaker sentiment from natural audio sources is a challenging problem. Generic methods for sentiment extraction generally use transcripts from a speech recognition system, and process the transcript using text-based sentiment classifiers. In this study, we show that this baseline system is sub-optimal for audio sentiment extraction. Alternatively, new architecture using keyword spotting (KWS) is proposed for sentiment detection. In the new architecture, a text-based sentiment classifier is utilized to automatically determine the most useful and discriminative sentiment-bearing keyword terms, which are then used as a term list for KWS. In order to obtain a compact yet discriminative sentiment term list, iterative feature optimization for maximum entropy sentiment model is proposed to reduce model complexity while maintaining effective classification accuracy. A new hybrid ME-KWS joint scoring methodology is developed to model both text and audio based parameters in a single integrated formulation. For evaluation, two new databases are developed for audio based sentiment detection, namely, YouTube sentiment database and another newly developed corpus called UT-Opinion Opinion audio archive. These databases contain naturalistic opinionated audio collected in real world conditions. The proposed solution is evaluated on audio obtained from videos in youtube.com and UT-Opinion corpus. Our experimental results show that the proposed KWS based system significantly outperforms the traditional ASR architecture in detecting sentiment for challenging practical tasks.

***Keywords:*** *Audio recognizer, Machine Learning, Sentiment Analysis, Maximum Entropy Text Sentiment Detection algorithm.*

## I. INTRODUCTION

In a large proportion of these videos, people depict their opinions about products, movies, social issues, political issues, etc. The capability of detecting the sentiment of the speaker in the video can serve two basic functions: (i) it can enhance the retrieval of the particular video in question, thereby, increasing its utility, and (ii) the combined sentiment of a large number of videos on a similar topic can help in establishing the general sentiment. It is important to note that automatic sentiment detection using text is a mature area of research, and significant attention has been given to product reviews, we focus our attention on dual sentiment detection in videos based on audio and text analysis. We focus on videos because the nature of speech in these videos is more natural and spontaneous which makes automatic sentiment processing challenging. In Particular, 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 conditions, (ii) foreign accents, (iii) spontaneous speech production, and (iv)diverse range of topics. Our approach towards sentiment extraction uses two main systems, namely, Automatic Speech Recognition (ASR) system and text-based sentiment extraction system. For text based sentiment extraction, we propose a new method that uses POS (Part-Of-Speech) tagging to extract text features and Maximum Entropy modelling to predict the polarity of the sentiments (positive or negative) using the text features. An important feature of our method is the ability to identify the individual contributions of the text features towards sentiment estimation. We evaluate the proposed sentiment estimation on both publically available text databases and videos. On the text datasets, This provides us with the capability of identifying key words/phrases within the video that carry important information. By indexing these key words/phrases, retrieval systems can enhance the ability of users to search for relevant information..

## II. LITERATURE REVIEW

[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

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 Cepstral coefficients (MFCC), Perceptual Linear Predictive (PLP), Filter Bank (FBANK), pitch, their first and second derivatives are used as frame-based features.

[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 In paper "Audio and Text based multimodal sentiment analysis using features extracted from selective regions and deep neural networks" An improved multimodal approach to detect the sentiment of products based on their multimodality natures (audio and text) is proposed. The basic goal is to classify the input data as either positive or negative sentiment. 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 fixed-length vector by pooling the activations of the last hidden layer over time.[2]

Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. Towards Real-time Speech Emotion Recognition using Deep Neural Networks. EPJ Data Science, 5(1):31, 2016: proposes a real-time SER system based on end-to-end deep learning. Namely, a Deep Neural Network (DNN) that recognizes emotions from a one second frame of raw speech spectrograms is presented and investigated. This is achievable due to a deep hierarchical architecture, data augmentation, and sensible regularization. Promising results are reported on two databases which are the ENTERFACE database and the Surrey Audio-Visual Expressed Emotion (SAVEE) database.[4] From paper "Machine Learning and Sentiment Analysis Approaches for the Analysis of Parliamentary Debates" the author seeks to establish the most appropriate mechanism for conducting sentiment analysis with respect to political debates; Firstly so as to predict their outcome and secondly to support a mechanism to provide for the visualisation of such debates in the context of further analysis. To this end two alternative approaches are considered, a classificationbased approach and a lexicon-based approach. In the context of the second approach both generic and domain specific sentiment lexicons are considered. Two techniques to generating domain-specific sentiment lexicons are also proposed: (i) direct generation and (ii) adaptation. The first was founded on the idea of generating a dedicated lexicon directly from labelled source data. The second approach was founded on the idea of using an existing general purpose lexicon and adapting this so that it becomes a specialised lexicon with respect to some domain. The operation of both the generic and domain specific sentiment lexicons are compared with the classification-based approach. The comparison between the potential sentiment mining approaches was conducted by predicting the attitude of individual debaters (speakers) in political debates (using a corpus of labelled political speeches extracted from political debate transcripts taken from the proceedings of the UK House of Commons). The reported comparison indicates that the attitude of speakers can be effectively predicted using sentiment mining. The author then goes on to propose a framework, the Debate Graph Extraction (DGE) framework, for extracting debate graphs from transcripts of political debates. The idea is to represent the structure of a debate as a graph with speakers as nodes and\exchanges" as links. Links between nodes were established according to the exchanges between the speeches. Nodes were labelled according to the \attitude" (sentiment) of the speakers, \positive" or \negative", using one of the three proposed sentiment mining approaches.

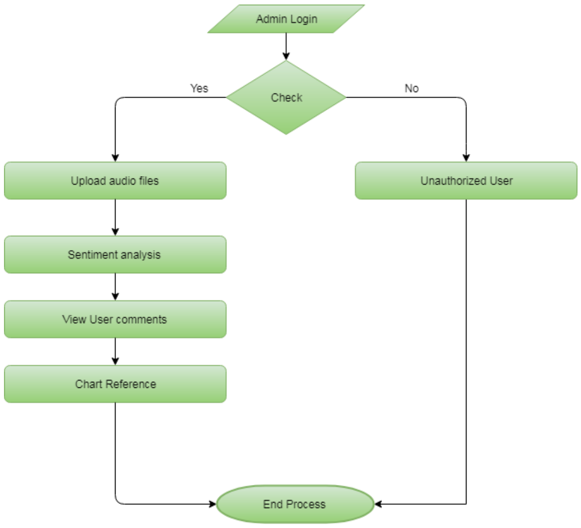


Fig.1. Admin Flow diagram of proposed system.

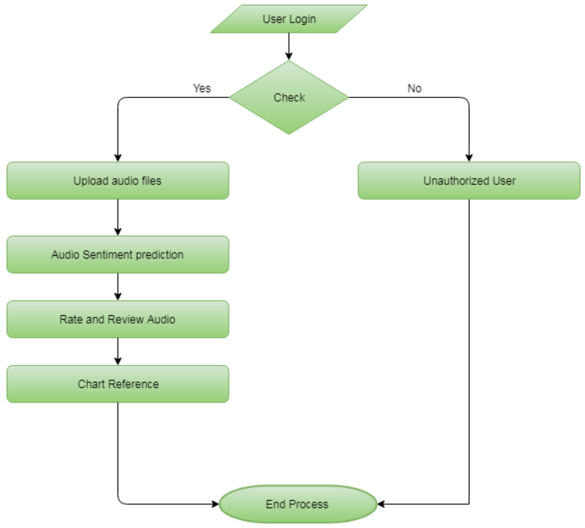


Fig.2. User Flow diagram of proposed system

First one is the offline text based sentiment model generation, the second is the ASR based sentient detection system forming our baseline system, and finally the third is a proposed system using audio Keyword Spotting (KWS) approach. Each block is explained in detail in subsequent sections. In reality, accurate sentiment detection generally relies on a small fraction of the speech recognition transcript, because sentiment bearing vocabulary tends to be sparse in spoken opinions. In other words, sentiment detection accuracy depends on being able to reliably detect and recognize a much focused vocabulary in the spoken comment audio stream. Therefore, keyword spotting (KWS) technology is expected to be better suited for sentiment detection, as opposed to full-transcript ASR.

## III. METHODOLOGY

### MAXIMUM ENTROPY TEXT SENTIMENT DETECTION

Approach: k-medoids and dynamic time warping A naive approach to clustering arcs could be to use a popular algorithm such as k-means with an Euclidean metric to measure the distance between two arcs. However, this is a poor approach for our problem for two reasons: Taking the mean of arcs can fail to find centroids that accurately represent the shapes in that cluster. a pathological example of when this occurs. The mean of the left two arcs has two peaks instead of one.

The Euclidean distance between two arcs doesn't necessarily reflect the similarity of their shapes. While the With these limitations in mind, we turn to k-medoids with dynamic time warping (DTW) as the distance function. K-medoids updates a medoid as the point that is the median distance to all other points in the cluster, while DTW is an effective measures the distance between two time series that may operate at different time scales.

A maximum entropy classifier starts off making the least assumption in terms of certainty about the underlying data distribution as illustrated in the three scenarios below

If we are trying to determine if a coin is fair we could start off assuming it is fair, that is both heads and tails are equally likely and revise our opinion as we perform more experiments. Same with a dice - we could start off assuming all six outcomes are equally likely as shown in figure below and then revise the assumption as we gather more data

If we are trying to find the distribution of heights of students in a school, and we have some prior knowledge of the spread of heights, then we can start off assuming the heights are distributed like a bell shape as shown in figure below (it would be too conservative to assume all heights are same - bell shape is an optimal start)

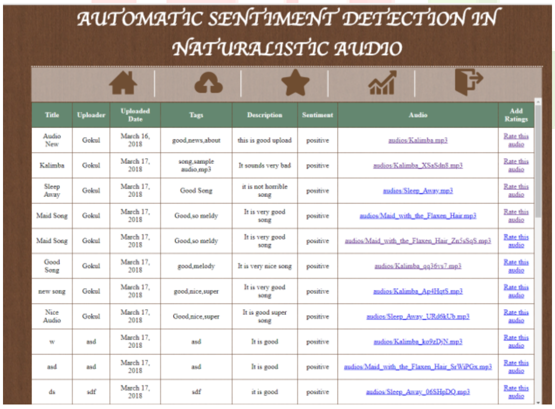
Lastly, if we are estimating the rate of radioactive decay of some element we have, and we have prior knowledge of the average rate of decay *(all positive values for outcomes)*, we can start off assuming a decay rate distribution like the one in figure below.

The key takeaway is in all three of the above cases, our starting assumptions are the optimal conservative assumptions in terms of uncertainty to get started. That is all three figures, specific to the three use cases, are the optimal highest uncertainty start points. Our uncertainty can only decrease, if at all, as we conduct more experiments to revise our beliefs.

## EXPERIMENTAL RESULTS AND DISCUSSION

### AUDIO SENTIMENT ON UPLOAD

Upload the audio with tags and description. Based on the description of the audio, we can sense that details as positive or negative. Based on the sentiment of description, we can analysis the sentiment of audio. This will detect the sentiment audio. Audios sentiment can be analyzed while admin or user uploading the audios with the tags and descriptions.



| Title | Uploader | Uploaded Date | Tags | Description | Sentiment | Audio | Add Ratings |
|---|---|---|---|---|---|---|---|
| Audio New | Gokul | March 16, 2018 | good,news,about | this is good upload | positive | audios/Kalimba.mp3 | Rate this audio |
| Kalimba | Gokul | March 17, 2018 | song,sample audio,mp3 | It sounds very bad | positive | audios/Kalimba_XSaSdn8.mp3 | Rate this audio |
| Sleep Away | Gokul | March 17, 2018 | Good Song | it is not horrible song | positive | audios/Sleep_Away.mp3 | Rate this audio |
| Maid Song | Gokul | March 17, 2018 | Good,so meldy | It is very good song | positive | audios/Maid_with_the_Flaxen_Hair.mp3 | Rate this audio |
| Maid Song | Gokul | March 17, 2018 | Good,so meldy | It is very good song | positive | audios/Maid_with_the_Flaxen_Hair_Zn5sSqS.mp3 | Rate this audio |
| Good Song | Gokul | March 17, 2018 | good,melody | It is very nice song | positive | audios/Kalimba_qq36vs7.mp3 | Rate this audio |
| new song | Gokul | March 17, 2018 | good,nice,super | It is very good song | positive | audios/Kalimba_Ap4HqtS.mp3 | Rate this audio |
| Nice Audio | Gokul | March 17, 2018 | Good,nice,super | It is good super song | positive | audios/Sleep_Away_URd6kUb.mp3 | Rate this audio |
| w | asd | March 17, 2018 | asd | It is good | positive | audios/Kalimba_ko9zDjN.mp3 | Rate this audio |
| asd | asd | March 17, 2018 | asd | It is good | positive | audios/Maid_with_the_Flaxen_Hair_SrWiPGx.mp3 | Rate this audio |
| ds | sdf | March 17, 2018 | sdf | it is good | positive | audios/Sleep_Away_06SHpDQ.mp3 | Rate this audio |

Fig. 3.User List

### USER REVIEWS AND RATINGS

User can give rating and review to the audio. And user can upload comments for user side. Admin view all user ratings and comments. Comments are analyzed whether it is positive or negative. Based on the review analysis this can be rate. Admin have the rights to remove the review completely from the database.

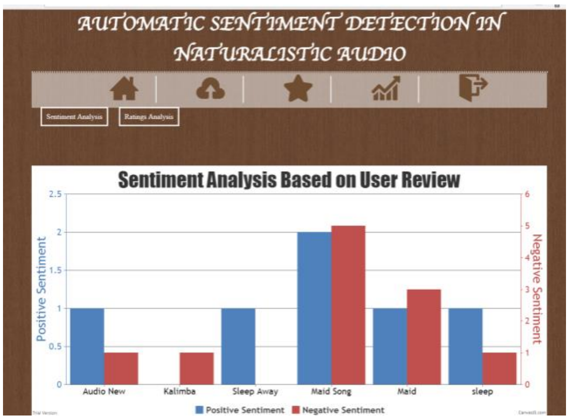
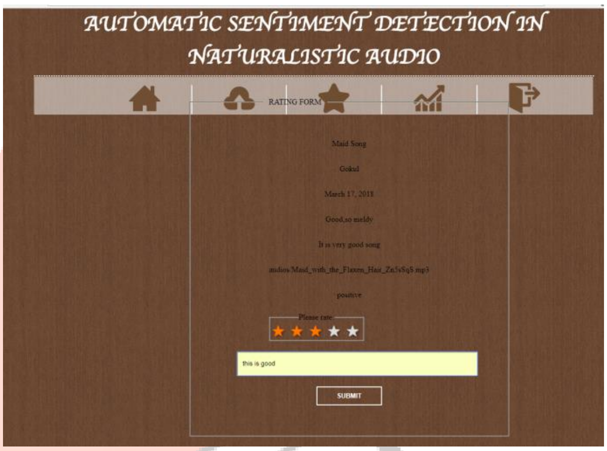


Fig. 5.User Ratings



RATING FORM

Maid Song

Gokul

March 17, 2018

Good,so meldy

It is very good song

audios/Maid_with_the_Flaxen_Hair_Zn5sSqS.mp3

positive

Please rate

★★★☆☆

this is good

SUBMIT

Fig. 6.Based on user Review

### GRAPH ANALYSIS

The Details of ratings and review and analyzed sentiment can be view as a chart for the convenience of analyzer to understand the details in proper manner. The chart in this system is plot as Column Chart, Pie Chart and Spline Chart. The charts shown in the both user and admin side as well.

### GRAPHICAL REPRESENTATION

This is graphical notation of the data given by the system. This phase of implementation will shows the effectiveness of the proposed system through pictorially in the order to better understand of proposed system.

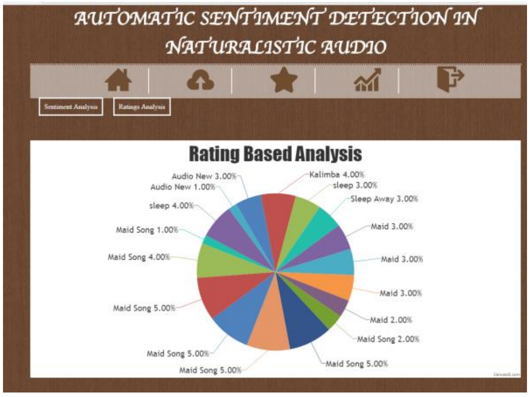


Fig. 7.Graphical Representation

## CONCLUSION

we have presented the evaluation of the proposed system on YouTube and UT-Opinion corpora. The new method has been compared to a baseline system that uses raw transcripts from ASR and feeds it to text based sentiment classifier. Our experimental results show that the new method outperforms the baseline system by reducing the error rate by 19% relative in YouTube, and 8% relative in UT-Opinion. While the new method improves upon audio based sentiment detection, there is room for further improvement. For example, addressing the traditional robustness problems of ASR (accent, noise, etc.) can have significant impact of performance. Another area of work could focus on using pure speech features to augment lexical information drawn for speech recognition to do speech sentiment detection.

## 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 speech 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.

[7] Jennifer S Lerner, Ye Li, Piercarlo Valdesolo, and Karim S Kassam. Emotion and decision making. Annual Review of Psychology, 66:799–823, 2015.

[8] Katherine L Milkman and Jonah Berger. The science of sharing and the sharing of science. Proceedings of the National Academy of Sciences, 111(Supplement 4):13642–13649, 2014.

[9] Andrew J Reagan, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. Towards Real-time Speech Emotion Recognition using Deep Neural Networks. EPJ Data Science, 5(1):31, 2016.

[10] Anna Rohrbach, Atousa Torabi, Marcus Rohrbach, Niket Tandon, Christopher Pal, Hugo Larochelle, Aaron Courville, and Bernt Schiele. Movie description. International Journal of Computer Vision, 123(1):94–120, 2017.

[11] Makarand Tapaswi, Yukun Zhu, Rainer Stiefelhagen, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. Movieqa: Understanding stories in movies through

question-answering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4631–4640, 2016.

[12] Atousa Torabi, Christopher Pal, Hugo Larochelle, and Aaron Courville. Using descriptive video services to create a large data source for video annotation research.
arXiv preprint arXiv:1503.01070, 2015.