**SENTIMENT ANALYSIS: SENTIMENT ANALYSIS ON AUDIO FOR HELP DESK**

Kashish Sharma  
Symbiosis Institute of Technology, Nagpur Campus, Symbiosis

International Deemed University, Pune, India   
[kashish.sharma.batch2021@sitnagpur.siu.edu.in](mailto:kashish.sharma.batch2021@sitnagpur.siu.edu.in)

Anmol Gupta  
Symbiosis Institute of Technology, Nagpur Campus, Symbiosis

International Deemed University, Pune, India   
[Anmol.gupta.batch2021@sitnagpur.siu.edu.in](mailto:Anmol.gupta.batch2021@sitnagpur.siu.edu.in)

Poorva Agrawal  
Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International Deemed University, Pune, India

Isha Sharma

Symbiosis Institute of Technology, Nagpur Campus, Symbiosis

International Deemed University, Pune, India   
[isha.sharma.batch2021@sitnagpur.siu.edu.in](mailto:isha.sharma.batch2021@sitnagpur.siu.edu.in)

***Abstract*.**

***Keywords*—**

***Introduction:***

In the digital era, the number of audio information has increased incredibly by raising number of online podcasts, social media, messengers and other communication tools. The diversity of sounds here transcends into a melody where human feelings, a variety of emotions and much more are encrypted in these sounds. Sentiment analysis as an area of NPL (natural language processing) with an aim of text and particularly to reveal the underlying meaning or opinion of individuals is a popular method. Nevertheless, with the soaring of audio content, the time has come to understand how to include sentiment analysis methods in the linguistic domain is essential.  
Describing moods from audio data creates an endless depth of betterment spanning marketing, users and feedback, to users' mental health and customized experiences. The scientists and experts in the domain decode why a person is aggressive/happy/sad/neutral out of what they say, and how that is infused in their spoken words, tone, and intonation.  
The purpose of this research document is to go more deeply into topic of emotion analysis for audio, presenting the necessary background, current issues, and evolving trends. By looking at how sentiment analysis has evolved from the word-based approaches to the multimodal techniques that can analyze both audio-visual content we will come to understand how sentiment analysis differs from traditional text-based analysis.

Recent discoveries in areas like machine learning algorithms, deep learning architectures, and signal processing techniques have enabled researches to go a long way towards mitigating these challenges and improving the accuracy and steadfastness of audio-based sentiment analysis systems.

Moreover, as we will show how sentiment analysis relates to practical areas and use cases such as the recommendation systems based on emotions, emotion-aware virtual assistants and application of affective computing in human-computer interactions. Our objective is to ensure that the discussed applications of sentiment analysis on audio are applied so that we can clearly realize the practical applicability and possible influence in the various realms.

In conclusion, this research paper aims to provide a comprehensive synthesis of the current state-of-the-art in sentiment analysis on audio, offering insights into the underlying methodologies, challenges, and applications. By fostering a deeper understanding of this burgeoning field, we hope to inspire further research and innovation, ultimately paving the way for more sophisticated and context-aware audio sentiment analysis systems.

***Literature Review:****Top of Form*

The analysis in audio has been one of the most stimulating topics in recent years. Despite the fact that audio analysis has emerged as an ultimate multi-functional tool that is now useful in many means of life - including business, helping customers, healthcare, and entertainment to name just a few. The aims of this literature review are to briefly present the actual level of development, challenges and ways forward in Auditory Information Theory.  
Breakthroughs in machine learning and deep learning in research has created the environment to accelerate development of algorithms which can extract and process audio profiles. Traditional techniques, the traditional combination of feature engineering and classical machine learning models have evolved into more powerful deep learning structures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and the likes of Transformer-based BERT.

During the pre-processing phase and feature extraction the key point to note is that these are the main challenges that emerge in audio analysis. Differently from text data that can be easily accessed from structured formats of data, the audio materials require pre-processing steps in order to cut audio for segmentation purposes, extract any feature of interest (e.g., MFCC, spectrogram) and noise reduction so as to improve the quality of the signal. Also, it is very hard to add contextual elements in audio sentiment analysis. Moreover, contextual elements aren’t the only factor in consideration, you also need consider the extra linguistic lectures like pitch, intonation, and rhythm. Numerous studies that can be referred to have resulted in audio fusion with text and video, designating multi-modal assets to extend the method of analysis. Emotional sensitivity was represented using an array of methods starting from early fusion (where the input-level features were combined) to late fusion (where models' prediction value was combined).  
The availability of datasets such as the RAVDESS and IEMOCAP datasets facilitates the progress of research as they offer an acceptable operation environment that is used to both train and evaluate different models and algorithms.  
Nevertheless, the existing problems of dynamic adaptation, noise robustness, and deep learning model interpretation impede better data precision and accuracy. To solve these problems, research team of specialists should work on, machine learning, signal processing, speech, and psychology.

In summary, it can be said that field of audio perception that is characterized by growth as well as perspective for real-world implementation. In the future, model development will be bypassed to achieve strong languages models able to deal with various inputs based on context-aware analysis. Approaches that

entail augmenting the domains of mental wellness and personalized suggestions depict critical

ventures. Entering through the endeavour of the challenges, emotional voice analysis makes it clearer

that we can draw mental states and behaviours broadly across the contexts.

Top of Form

Bottom of Form

**Sentiment** **Model**

**Pre-processing**

**Output as text**

**Audio**

**Parsing**

**Positive**

**Neutral**

**Negative**

*Fig1.1.Seentiment Analysis on text*

***Dataset:***

The RAVDESS dataset is not limited to the major emotions, that voice actors convey in many settings and situations. Each audio file is approximately 3-5 seconds long and it is labelled with one of eight emotional states: calm, happy, sad, angry, fearful, surprised, disgusted, or neutral. Taken together, the 24 voice actors who participated in the research offer a vast range of vocal parameters - including accents as well as emotional intensities.

***Proposed Methodology:***

**Mel-Frequency Cepstral Coefficients(MFCC) :**

MFCCs are one of the commonly used types of attributes in speech and audio processing, such as for speech recognition, speaker identification, or sentiment analysis. The MFCCs signify the entire energy spectrum of sound over a short period that make it involve both frequency and time factors in the sound recording. Mel-Frequency Cepstral Coefficients (MFCCs) extraction involves several steps. At first, audio signal is divided into short overlapping frames to capture temporal variations, followed by windowing using functions like Hamming window to reduce spectral leakage. This will be followed by applying Fast Fourier Transform (FFT) on each windowed frame to compute power spectrum, which efficiently represents frequency content within a frame. Subsequently, Mel filter bank analysis is employed where overlapping triangular filters spaced on the Mel scale are applied to FFT obtained power spectrum that emphasizes important frequency components while reducing dimensionality. Afterward, Discrete Fourier Transform (DFT) converts filtered spectrum back to time domain emphasizing perceptually relevant features. In the e­nd, the MFCCs we acquire act as a ne­at summary that grabs both the time-relate­d and spectral qualities of the signal. This is pe­rfect for gauging sentiments, spe­cifically when merged with machine­ learning methodologies like­ Support Vector Machines (SVM). It provides a pre­cise sentiment cate­gorizing mechanism in audio data.

**Grid Search:**

Machine Learning Algorithms use Grid search, a structured approach to fine-tuning the hyperparameters mainly used in sentiment analysis exercises involving the application of MFCC features and SVM. The performance of SVM models is significantly affected by hyperparameters such as the regularization parameter (C), kernel function, and kernel-specific parameters. The choice of kernel function determines how well the model can capture detailed patterns on the data, while the regularization parameter (C) balances between margin size and classification error. Additionally, behavior for specific kernels is tuned by these kernel-specificparameters. Grid search conducts a systematic exploration

through predetermined combinations of hyperparameters. Various settings of hyperparameter are tried out when training SVM models iteratively on training data and their performances evaluated on validating set. This process runs until all combinations have been tested and the best performing combination for validation set identified as a result. Validation performance based on different hyperparameters is computed by means of grid search methodology. Consequently, its efficiency in real world sentiment analysis scenarios is measured by evaluating generalization performance using a separate test set. This thorough method guarantees that the SVM model is fine tuned for analysing sentiments enabling classification of emotions, in recordings.

**Support Vector Machine(SVM):**

Supervised machine learning algorithm called Support Vector Machine (SVM) is used for classiﬁcation  and regression tasks. The primary goal of SVM is to find the optimal hyperplane that excellently separates data points, belonging to different classes in the feature space, so as to maximize the margin between the hyperplane and nearest data points (support vectors). With the kernel trick, which allows it to implicitly map input features into higher-dimensional spaces, SVM works well for both linearly and non-linearly separable data. In order to provides robust classification and regression predictions by finding a hyperplane with maximum margin, SVM has wide application in different machine learning scenarios as a widely used and versatile algorithm.

**Bottom of Form**

Input Data

Audio

Pre-Processing

Feature Extraction

MFCC,Labels

Feature Selection

Tuned Hyperparameter

Grid Search

Support Vector Machine(SVM)

*Fig.1.2.Methodology*

***Results:***

***Comparative Results:***

***Findings:***

***Conclusion:***

In conclusion, using such customer sentiment analysis is a very crucial step for the organizations that aim to grow the customer base and at the same time strengthen their loyalties. This paper proposes a new sentiment analysis method that puts a bimodal sentiment feature space into sound data. By implementing intensive experiments using RAVDESS dataset, which is the special designed set for audio sentiment analysis,

Our research shows that unlike other implementations which mostly use text, the current attempts at measuring customer sentiment through audio representations, particularly in support center environments are more effective. The outcome of this sentiment analysis model demonstrated high effectiveness in the sense to locate and interpret the sentiment from the audio data; imparting the fact that it can be used to improve the customer relationship management.

As a result of our research, the importance of using audio modalities including how to do that in sentiment analysis frameworks, especially in places where auditory clues are a vital component such as customer support interactions, is now very well known. Leveraging audio data, companies are able to unearth the hidden pitfalls of customer attitude, which in turn helps them to provide adjusted services and feedback to till customers’ requests are met.

Hence our research shows the key role of utilizing various modes of data input and employing high-tech data analytical tools for discovering useful information from the customer traces. Going ahead, more investigation and the improvement of audio-induced sentiment analysis approach could be the foundations of social relationship management and customer devotion plugs.

***References:***

[1]. Multi-Modal Sentiment Analysis Using Text and Audio for Customer Support Centers Hardik Srivastava

SRM Institute of Science and Technology.

[2].Machine Learning-Based Sentiment Analysis of Incoming Calls on Helpdesk Suhas Chavan

Vishwakarma University.

[3]. Audio and Text Sentiment Analysis of Radio

Broadcasts Naman Dhariwal1, Sri Chander Akunuri1

, Shivama1 and Sharmila Banu K21Department of Software Systems, Vellore Institute of Technology, Vellore-632014, India.

[4]. P. Mishra, R. Rajnish and P. Kumar,” Sentiment analysis of Twitter data: Case study on digital India,” 2016 International Conference on

Information Technology (InCITe) - The Next Generation IT Summit onthe Theme - Internet of Things: Connect your Worlds, Noida, India, 2016,

[5]. R. R. Sehgal, S. Agarwal and G. Raj,” Interactive Voice Response using Sentiment Analysis in Automatic Speech Recognition Systems,”2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), Paris, France, 2018.

[6]. Y. Chen and Z. Zhang,” Research on text sentiment analysis based on CNNs and SVM,” 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, China, 2018.

[7]. Pang, B., & Lee, L. (2004, July). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd annual meeting on Association for Computational

Linguistics (p. 271). Association for Computational Linguistics.

[8]. J. S. Vimali and S. Murugan,” A Text Based Sentiment Analysis Model using Bi-directional LSTM Networks,” 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India.

[9]. M. A. Hossan, S. Memon and M. A. Gregory,” A novel approach for MFCC feature extraction,” 2010 4th International Conference on Signal Processing and Communication Systems, Gold Coast, QLD, Australia.

[10]. B. M. A. Tahayna, R. K. Ayyasamy and R. Akbar,” Automatic Sentiment Annotation of Idiomatic Expressions for Sentiment Analysis Task,” in IEEE Access,