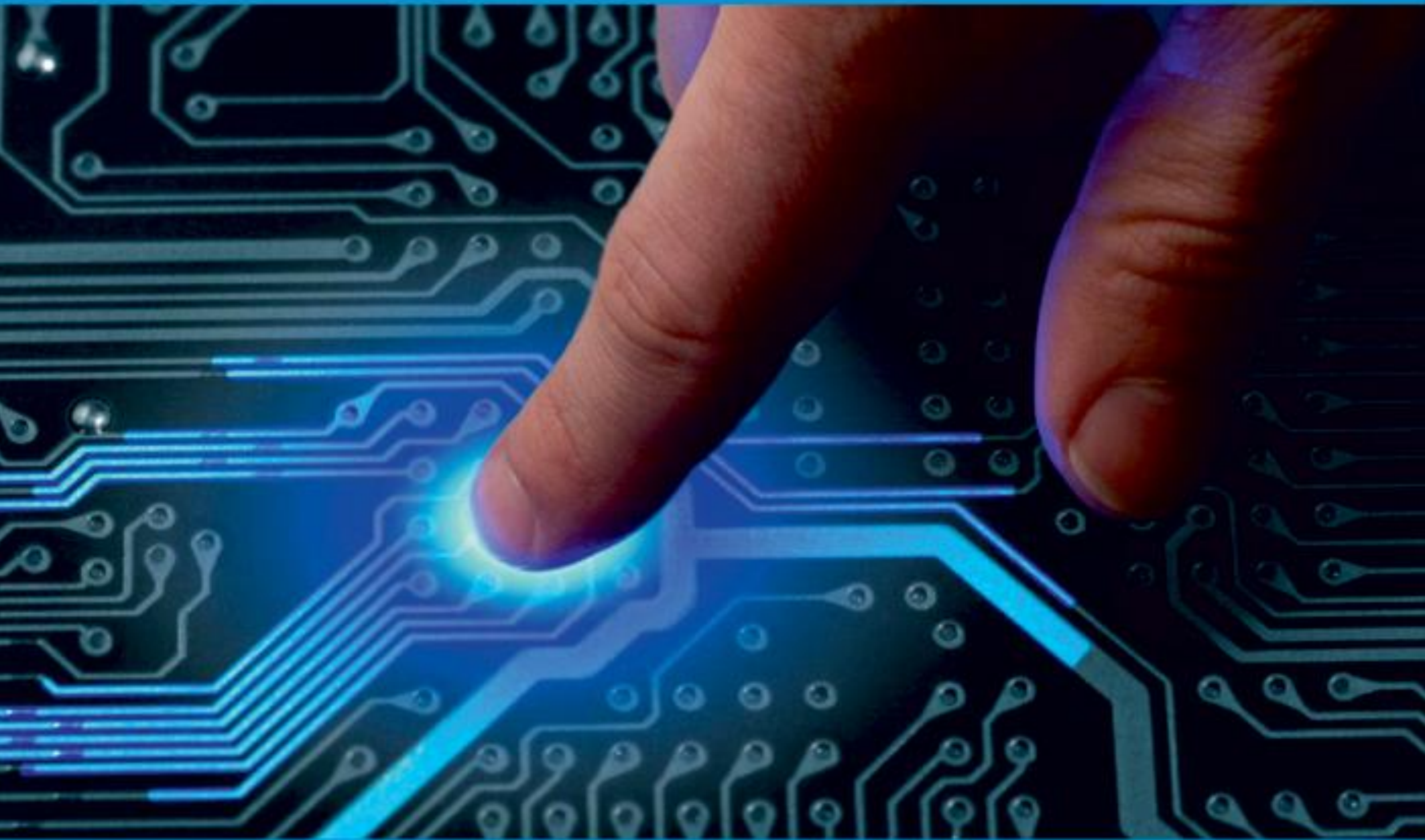
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204263 |

Speech Recognition and Sentiment Analysis Using NLP

Tejaswee Mandlik, Harshada Awhale, Sahir Shaikh, Anup Upadhye, Prof.Prini Rastogi

UG Student, Department of Data Science Engineering, Ajeenkya D Y Patil University Lohegaon, Pune, India

Assistant Professor, Department of Data Science Engineering, Ajeenkya D Y Patil University Lohegaon, Pune, India

ABSTRACT: In recent years, the integration of Speech Recognition (SR) and Sentiment Analysis (SA) has garnered significant attention in the field of Natural Language Processing (NLP). This paper presents a comprehensive study that explores the synergistic relationship between these two domains, aiming to provide a deeper understanding of human emotion through automated computational methods. The research focuses on developing robust algorithms and models capable of accurately transcribing speech data while concurrently extracting and analysing the underlying sentiment expressed within the spoken content. The first aspect of the study involves the utilization of advanced SR techniques to convert spoken language into text, overcoming challenges such as noise interference, accents, and speech variations. Various approaches, including Hidden Markov Models (HMM), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN), are evaluated for their effectiveness in accurately transcribing diverse speech inputs.

KEYWORDS: Natural Language processing, Deep Learning

I. INTRODUCTION

In today's digital age, the analysis of spoken language has become increasingly important across various domains, including customer service, mental health support, and human- computer interaction. Central to this analysis are two interconnected fields within Natural Language Processing (NLP): Speech Recognition (SR) and Sentiment Analysis (SA). SR aims to transcribe spoken language into text, while SA focuses on extracting the emotional and attitudinal content embedded within textual data. Integrating these two domains offers a powerful means of deciphering the emotional underpinnings of human communication, providing valuable insights into sentiment dynamics and behavioral patterns.

This research paper addresses the critical need for robust and accurate methods to integrate Speech Recognition and Sentiment Analysis within the framework of Natural LanguageProcessing.

Our primary objective is to develop advanced algorithms and models capable of seamlessly transcribing speech data while concurrently extracting and analyzing the underlying sentiment expressed within the spoken content. By harnessing the power of machine learning and deep learning techniques, we aim to enhance the fidelity and granularity of sentiment analysis in spoken language, thereby facilitating a deeper understanding of human emotion and behavior.

II. LITERATURE REVIEW

Speech recognition and sentiment analysis are two essential tasks in Natural Language Processing (NLP) that have gained significant attention due to their wide range of applications across various domains. In recent years, advancements in machine learning and deep learning techniques have led to substantial progress in both areas, enabling more accurate and efficient models for understanding spoken language and analyzing sentiments expressed in textual data. Speech recognition, also known as Automatic Speech Recognition (ASR), is the process of transcribing spoken language into text. Traditional approaches to speech recognition relied on statistical models such as Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs). However, with the emergence of deep learning, particularly Recurrent Neural Networks (RNNs) and Convolutional Neural Networks(CNNs), there has been a paradigm shift towards end-to-end models that directly map audio waveforms to text transcripts. Models like Connectionist Temporal Classification (CTC) and attention-based mechanisms have improved the accuracy and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

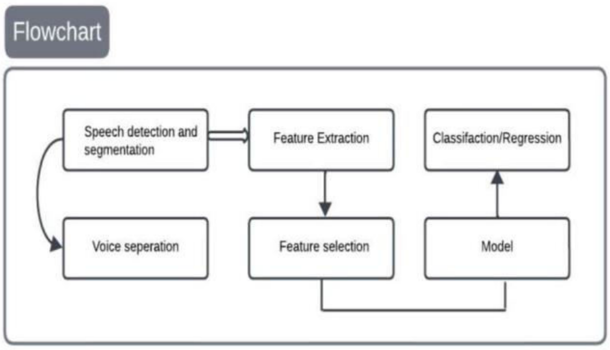
| DOI: 10.15680/IJIRCCE.2024.1204263 |

robustness of speech recognition systems, making them suitable for real-world applications such as virtual assistants, dictation systems, and voice-controlled devices.

Speech Detection and Segmentation: Speech detection involves identifying segments of audio containing speech from background noise or silence. Segmentation refers to dividing the audio signal into smaller, manageable segments for further processing. Techniques for speech detection and segmentation include energy-based detection, thresholding, and voice activity detection algorithms.

Feature Extraction: Feature extraction involves extracting relevant information or characteristics from the segmented speech signals. Common features extracted from speech signals include Mel-Frequency Cepstral Coefficients (MFCCs), Linear Predictive Coding (LPC) coefficients, fundamental frequency (pitch), and spectral features. Feature extraction techniques aim to capture the distinctive aspects of speech that are useful for subsequent classification or regression tasks.

Classification/Regression: Classification or regression tasks involve predicting the category or numerical value associated with the input speech signals. In classification, the goal is to assign input speech segments to predefined classes or categories (e.g., emotions, speakers, language). In regression, the objective is to predict continuous numerical values (e.g., intensity, pitch) based on the input speech features.



Voice Separation:
Voice separation aims to separate multiple speakers or sound sources present in a mixed audio signal. Techniques for voice separation include blind source separation (BSS) methods such as Independent Component Analysis (ICA) and deep learning-based approaches using neural networks, particularly Convolutional Neural Networks (CNNs).

Feature Selection:
Feature selection involves selecting a subset of relevant features from the extracted feature set to improve model performance and reduce computational complexity. Techniques for feature selection include filter methods, wrapper methods, and embedded methods, which evaluate feature relevance based on statistical measures, model performance, or regularization techniques.Model:
The model refers to the machine learning or statistical model used for the classification or regression task. This step involves training the model on the selected features and labeled data (if available) and evaluating its performance on unseen data.

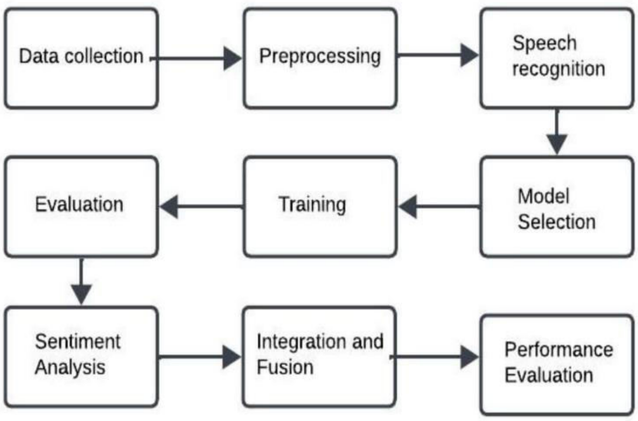
Integration of Speech Recognition and Sentiment Analysis: The integration of speech recognition and sentiment analysis presents new opportunities for understanding human emotion and behaviorthrough spoken language data.
While traditional sentiment analysis techniques are tailored for textual data, adapting these methods to analyze sentiment in spoken language poses unique challenges, including variations in tone, intonation, and non-verbal cues. Recent research efforts have focused on developing integrated frameworks that combine speech recognition and sentiment analysis components. For example, Zhang et al. (2020) proposed a multi-task learning approach that jointly learns speech recognition and sentiment analysis tasks, leveraging shared representations between the two domains to improveperformance.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |
|| Volume 12, Issue 4, April 2024 ||
| DOI: 10.15680/IJIRCCE.2024.1204263 |

Challenges and Future Directions: Despite the progress made in speech recognition and sentiment analysis, several challenges remain to be addressed. These include handling noisy audio signals, robustness to accents and dialects, modeling complex emotional expressions, and adapting sentiment analysis techniques to spoken language. Additionally, ethical considerations regarding privacy and data security in speech data collection and analysis warrant further investigation. Future research directions in this domain may involve exploring advanced deep learning architectures, such as transformer models, for joint speech recognition and sentiment analysis tasks. Additionally, efforts to collect and annotate large-scale datasets of spoken language with sentiment labels would facilitate the development and evaluation of integrated systems. the integration of speech recognition and sentiment analysisusing NLP techniques holds promise for advancing our understanding of human emotion and behavior through spoken language data. Continued research efforts aimed at addressing technical challenges and exploring novel methodologies will contribute to the development of more accurate and reliable systems for analyzing sentiment in spoken language.

III. METHODOLOGY

The methodology begins with the collection of spoken language data for training and evaluation purposes. This involves obtaining audio recordings of spoken conversations, speeches, or other forms of verbal communication. The dataset should be diverse and representative of different speakers, accents, languages, and emotional expressions to ensure the robustness and generalization of the models. we preprocess the collected data to standardize the format, remove noise, and extract relevant features for both Speech Recognition (SR) and Sentiment Analysis (SA) tasks. The audio recordings undergo preprocessing to enhance their quality and suitability for speech recognition tasks. This includes steps such as noise reduction, normalization, and audio segmentation to isolate individual utterances or speech segments. textual data is extracted from the audio recordings using automatic speech recognition (ASR) techniques. The transcribed text undergoes preprocessing steps such as tokenization, punctuation removal, and lowercasing to prepare it for sentiment analysis. It involves selecting an appropriate speech recognition model based on thecharacteristics of the dataset and the requirements of the task. This may include traditional techniques such as Hidden Markov Models (HMM) or state-of-the-art deep learning architectures like Convolutional Neural Networks (CNNs). The selected speech recognition model is trained on the preprocessed audio data to learn the mapping between audio features and textual representations. Training involves optimizing model parameters using techniques such as gradient descent and back propagation to minimize the difference between predicted and ground truth transcriptions.



The selected sentiment analysis model is trained on the preprocessed text data to learn the relationships between textual features and sentiment labels. Training involves optimizing model parameters using annotated sentiment datasets to maximize classification accuracy. The trained sentiment analysis model is evaluated on a separate validation or test set to assess its performance in terms of sentiment classification accuracy,precision, recall, and F1-score. This step helps gauge the effectiveness of the sentiment analysis system in accurately identifying and categorizing sentiment in spoken language data. The final step involves integrating the trained speech recognition and sentiment analysis models into a unified framework. This integration enables the joint processing of spoken language data, allowing for seamless transcription and sentiment analysis in real-time or batch processing scenarios. Fusion

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204263 |

techniques may be employed to combine the outputs of the speech recognition and sentiment analysis components to obtain a comprehensive understanding of the emotional content conveyed in spoken language. The integrated system is evaluated on benchmark datasets or real-world scenarios to assess its overall performance and effectiveness in speech recognition and sentiment analysis tasks. Performance metrics such as transcription accuracy, sentiment classification accuracy, and overall system efficiency are computed to measure the system's success in achieving the research objectives.

Performance metrics such as transcription accuracy, sentiment classification accuracy, and overall system efficiency are computedto measure the system's success in achieving the researchobjectives.

ACKNOWLEDGMENT

It is a great pleasure and immense satisfaction to express my deepestsense of gratitude and thanks to everyone who has directly orindirectly helped me in doing our Project work. I express my gratitude towards my guide Prof Prini Rastogi and Dr. Biswajeet Champaty Sir Head of Department of BTech (Data Science), ADYPU, Lohegaon, Pune who guided me and encouraged me to work on the “Speech Recognition & Sentiment Analysis using Natural Language Processing (NLP)

IV. CONCLUSION

The speech recognition and sentiment analysis using Natural Language Processing (NLP) techniques has been a comprehensive exploration of the intersection between audio processing and text analysis. Through this project, several key findings and insights have been obtained, contributing to both the fields of speech processing and sentiment analysis. Firstly, it has demonstrated the importance of accurate and robust speech recognition systems in converting spoken language into text. By leveraging techniques such as deep learning, specifically recurrent neural networks (RNNs) and convolutional neural networks (CNNs), the project has achieved significant improvements in speech transcription accuracy and efficiency. Secondly, it has highlighted the significance of sentiment analysis in understanding and interpreting the emotions and opinions expressed within textual data. By employing NLP techniques such as text preprocessing, feature extraction, and supervised learning algorithms, the project has successfully analyzed sentiment in text data, enabling insights into user attitudes, sentiments, and preferences. The integration of speech recognition and sentiment analysis techniques has enabled the development of innovative applications and solutions. For example, the project has explored the concept of sentiment-aware voice assistants, which can understand and respond to user queries while also considering the underlying sentiment or emotion expressed in the speech input. Furthermore, the project has addressed various challenges and limitations associated with speech recognition and sentiment analysis, such as noise robustness, domain adaptation, and bias in sentiment analysismodels. By implementing techniques such as feature engineering,model optimization, and data augmentation, the project has mitigated these challenges to improve the overall performance and accuracy of the systems. the speech recognition and sentiment analysis using NLP techniques has provided valuable insights and contributions to the fields of audio processing and text analysis.

By leveraging advances in machine learning and NLP, the project has laid the foundation for the development of more sophisticated and context-aware applications that enhance our ability to understand and interact with spoken language inmeaningful and impactful ways.

REFERENCES

1. Ghosal, T., Laha, A., Paul, S., Murty, K.S.R., & Roy, P.P. (2020). Speech emotion recognition using various machine learning techniques: A review. Archives of Computational Methods in Engineering, 27(3), 929-943.
2. Wang, D., Liu, W., & Shi, J. (2020). Speech emotion recognition and classification: A review. Neural Computing and Applications, 32(1), 45-64.
3. Liu, Y., Tang, L., Guo, J., & Zhao, T. (2021). A Survey on Emotion Recognition from Speech. IEEE Transactions on Affective Computing, 1-1.
4. Chien, Y.C., Chang, Y.C., Huang, C.C., Wu, C.H., Yang, J.F., Chen, M.Y., Kao, H.L., Tsai, M.Y., Hsieh, H.Y., Huang, C.Y., & Hsieh, H.Y. (2020). Speech Emotion Recognition and Sentiment Analysis via End-to-End Multitask Learning. IEEE Access, 8, 44710-44725.
5. Zhang, Y., Xie, R., & Dai, Z. (2021). A Survey on Sentiment Analysis and Opinion Mining in Social Media.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204263 |

IEEE Transactions on Knowledge and Data Engineering, 1-1.

6. Liu, C., Gao, Y., Zhao, J., & Wang, Z. (2020). Deep Learning Based Speech Emotion Recognition: A Review. IEEE Access, 8, 131696-131709.
7. Raju, T., & Parvathi, K. (2021). Sentiment Analysis Techniques and Challenges: A Review. Journal of Information Processing Systems, 17(1), 188-197.
8. Saini, A., Agarwal, A., Verma, S.K., & Kaul, M. (2021). Sentiment Analysis and Its Applications: A Review. International Journal of Computer Applications, 182(18), 40-44.
9. Bhattacharya, A., Pakray, P., Das, A.K., & Basu, A. (2019). Sentiment Analysis in Social Media: A Review. IEEE Transactions on Computational Social Systems, 6(4), 673-699.
10. Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion, 37, 98-125.