Sentiment Prediction Using Voice

Jaya Vishwakarma[1], Keshav Sharma[2]

1,2 Students, Department of Computer Science and Engineering, Babu Banarasi Das Northern India Institute of Technology

Dr. Anurag Shrivastava

Head of Department of Computer Science & Engineering, Babu Banarasi Das Northern India Institute of Technology

---***---

Abstract - The project ‘Sentiment Analysis using Voice’ will help in identifying the fraud calls like scammers asking for bank details, personal details of less educated people ,who believe these frauds easily and believe they are actually the officials from the agency ,and got scammed by them .The limitations to this can be noise and disturbance in audio data and background noises to minimize the impact of these errors we are working on a model to provide a viable solution. The project can also be used in detecting negative emotions and harmful thoughts on more research and on including some additional features. The project will come in handy when used on calls for real time analysis of call, for detecting the call is fraud or valid on time. It will start from basic after learning process , understanding sound and it’s features in machine from different sources , and how to handle sound data , how to preprocess it and how to build model for it. The final step involves deployment and testing of model on real voice data provided by one of our team members.

***Key Words*:** sentiment, voice, light, dataset, machine learning, testing.

1.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..

2. LITERATURE REVIEW

- From paper” 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 Cepstral coefficients (MFCC), Perceptual Linear Predictive (PLP), Filter Bank (FBANK), pitch, their first and second derivatives are used as frame-based features.
- 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.
- The paper “Towards Real-time Speech Emotion Recognition using Deep Neural Networks” proposes a real-time SER system based on endto-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.

- 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.
- The paper "Towards Real-time Speech Emotion Recognition using Deep Neural Networks" proposes a real-time SER system based on endto-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.

- In paper" Sentiment extraction from natural audio streams" a system for automatic sentiment detection in natural audio streams such as those found in YouTube is proposed. The proposed technique uses POS (part of speech) tagging and Maximum Entropy modelling (ME) to develop a text-based sentiment detection model. Additionally, we propose attuning technique which dramatically reduces the number of model parameters in ME while retaining classification capability. Finally, using decoded ASR (automatic speech recognition) transcripts and the ME sentiment model, the proposed system is able to estimate the sentiment in the YouTube video. In our experimental evaluation, we obtain encouraging classification accuracy given the challenging nature of the data. Our results show that it is possible to perform sentiment analysis on natural spontaneous speech data despite poor WER (word error rates).

3. PROBLEM STATEMENT

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.

4. OBJECTIVES OF THE STUDY

- Sentiment classification is a way to analyze the subjective information in the text and then mine the opinion „Sentiment analysis is the procedure by which information is extracted from the opinions, appraisals and emotions of people in regards to entities, events and their attributes . In decision making, the opinions of others have a significant effect on customers ease, making choices with regards to online shopping, choosing events, products, entities.

- Categorization or classification of opinion sentiment into- • Positive Negative

5.EXISTING SYSTEM

Research on speech sentiment analysis and emotion recognition is not new. Previous studies have focused on building feature extraction methods or classifiers for general speech processing or specific cases. In this section, we highlight the previous research on sentiment analysis and emotion recognition from speech and emphasize the key differences in our approach in this study compared to those of the previous studies.

Bertero et al. [14] performed speech emotion and sentiment recognition for interactive dialogue systems. The authors annotated the TED-LIUM release 2 corpus for emotion recognition and employed Movie Review and Twitter corpora for sentiment analysis. Those authors utilized raw speech for emotion recognition and word2vec for sentiment analysis. It is clear that although the authors proposed to undertake both sentiment analysis and emotion recognition, they used two different models for each task (using different datasets). For emotion recognition, the authors evaluated the use of the SVM and CNN methods. For sentiment analysis, the authors evaluated the use of the CNN and LIWC (Linguistic Inquiry and Word Count) methods.

The authors proposed a method to extract features from visual and textual modalities using a deep CNN for emotion recognition and sentiment analysis. The method, which is called multiple kernel learning, is a feature selection method to combine data from different modalities effectively. The multimodal data consisted of audio, video, and text. The proposed method with feature selection slightly improved the performance of multimodal fusion without feature selection. The gap in the performance of multimodal fusion over any single modality was large, highlighting the benefit of the use of multimodal fusion over the feature selection method.

6. PROPOSED SYSTEM

- Transcribe call recordings and determine the sentiment of callers-

- Improving customer experience is a key objective for contact centers. Generating data based on customers' experiences is, therefore, beneficial. However, obtaining this data is often a challenge. Contact centers are now using Speechmatics' any-context speech recognition engine to generate accurate transcripts from calls. With voice data transformed into a textbased format, it can then be used for sentiment analytics as part of the wider offering, workflow, or solution.

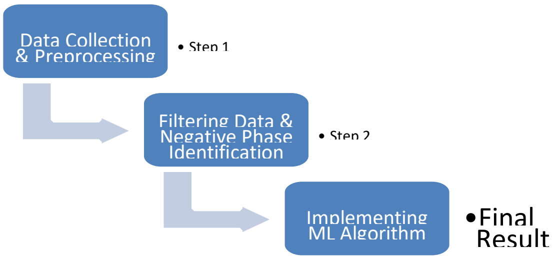
Data Collection & Preprocessing
• Step 1
Filtering Data & Negative Phase Identification
• Step 2
Implementing ML Algorithm
•Final Result

Figure 6.1: Planning of work

7. PLANNING OF WORK

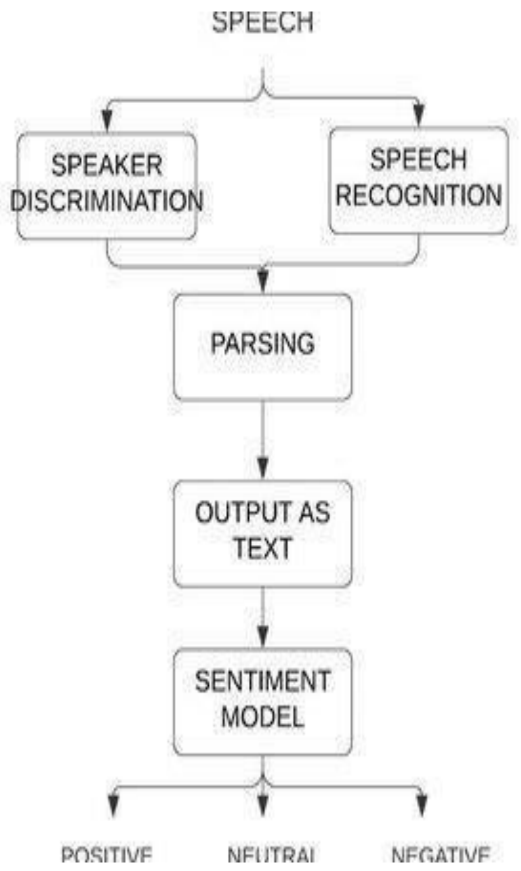

Figure 6.2: Architecture Design

8. FUTURE SCOPE

- Finding an efficient way to connect model to web app.
- Training model on our voices.
- Improving model

9. CONCLUSION

Through this project, we showed how we can leverage Machine learning to obtain the underlying emotion from speech audio data and some insights on the human expression of emotion through voice. This system can be employed in a variety of setups like Call Centre for complaints or marketing, in voice-based virtual assistants or chatbots, in linguistic research, etc.

REFERENCES

1. Sentiment Analysis of speaker specific speech data

   Author: Maghilnan S, Rajesh Kumar M,
   Senior IEEE, Member School of Electronic Engineering
   VIT University Tamil Nadu,India.

   □

2. Sentiment Analysis of Speech

   Author : Aishwarya Murarka , Kajal
   Shivarkar , Sneha, Vani Gupta ,Prof.Lata
   Sankpal Student,
   Department of Computer Engineering,
   Sinhgad Academy of Engineering, Pune,
   India

3. Real-Time Speech Emotion Recognition Using a Pre-trained Image Classification Network:
   Effects of Bandwidth Reduction and Companding-

   Author : Lech M, Stolar M, Best C and Bolia R (2020).

4. Pre-trained Deep Convolution Neural Network Model With Attention for Speech Emotion Recognition-

   Author : Zhang H, Gou R, Shang J,
   Shen F, Wu Y and Dai G

5. Stacking machine learning models for speech sentiment analysis-

   Author : Zadeh A, Liang PP, Poria S, Vij P, Cambria
   E and Morency.