Personal Journal with Emotional Intelligence and Sentiment Tracking

Vysakh Jayan, Nikhil Anilkumar, Augustine B, Amal Saji Varghese, Poorna B R

Abstract— Emotional Intelligence is the ability to analyze and react according to emotions of a person. The systems having emotional intelligence can have many potential applications including highly personalized devices and psychological support systems. The emotional quotient of a person is highly related to quality of life, productivity, social instincts etc. Thus, it is a highly relevant field and will be able to provide multitude of benefits to the user and the society in general. The objective of this system is to create a personal journal that analyzes and reacts according to the classified emotions of the user, while generating information from the emotional statistics. The system will provide an aesthetic UI for user interaction while running the emotion classification system and conversational intelligence system to react according to user's emotions. The system will also save the diary entries of the user and provides psychological benefits to the user.

Index Terms— Emotional Intelligence, Speech Emotion Recognition (SER), MFCC (Mel Frequency Cepstral Coefficient), Web Application, UI (User Interface), Diary Management, NLTK (Natural Language ToolKit), Django Framework, LSTM (Long Short-Term Memory), Accuracy

1 Introduction

The role of emotions in the life of a person is crucial. Emotions are the driving factor behind the decisions a person take in their life. Thus, analyzing emotions could yield important information regarding a person's mental health, interests, likes and dislikes. This information could be used to design user-centric products and services. Emotional well-being is crucial to the quality of life of a person. Thus, a system sensitive to human emotions can prove to be highly contributive to the society. The proposed system is a digital twin of a personal journal with emotional intelligence capabilities. A personal journal holds sensitive and crucial information about the person. Thus, it is the best source to generate the most credible information regarding their emotional statistics. Additionally, providing emotional support through such a system is highly effective. Essentially, the system aims at being a friendly and emotionally reactive personality for the user to interact while taking the hassle off of maintaining a personal journal.

- *Vysakh Jayan is currently pursuing bachelors degree program in computer science engineering in Mar Baselios College of Engineering and Technology, Kerala, India, PH-00919995721771. E-mail: vysakhjayandg@gmail.com*
- *Nikhil Anilkumar is currently pursuing bachelors degree program in computer science engineering in Mar Baselios College of Engineering and Technology, Kerala, India, PH-00919567155754. E-mail: nikhil.anilkumar.98@gmail.com*
- *Augustine B is currently pursuing bachelors degree program in computer science engineering in Mar Baselios College of Engineering and Technology, Kerala, India, PH-00919846521591. E-mail: augustineb789@gmail.com*
- *Amal Saji Varghese is currently pursuing bachelors degree program in computer science engineering in Mar Baselios College of Engineering and Technology, Kerala, India, PH-00919605570732. E-mail: amalsajivarghese8@gmail.com*
- *Poorna B R, Assistant Professor, Computer Science Engineering Department in Mar Baselios College of Engineering and Technology, Kerala, India, PH-00919446356114. E-mail: poorna.br@mbcet.ac.in*

2 Literature Review

Various techniques for speech emotion recognition systems are being conceptualized by engineers. Some of the techniques with their properties of feature matching and feature extraction provide feasible results. In most methodologies, it is found that MFCC is the widely used audio feature component used for feature extraction.

Some other proposed systems use speech emotion recognition systems using RNN and SVM models to classify basic human emotions. These systems mainly extracted two sets of speech signal features, namely MS and MFCC, from the Spanish and Berlin databases, and this combination of the selected attributes resulted in moderately accurate classifications [4]. Only the most dominant features are for training. Many feature extraction methods prove that not all data is important for accurate classification. Speech features based 2D CNN model provides better accuracy relative to state-of-the-art results, which further improves when combined with text. The combined Spectrogram-MFCC model results in an overall emotion detection accuracy of 73.1%, an almost 4% improvement to the existing state-of-the-art methods. Better results are observed when speech features are used along with speech transcriptions. The combined Spectrogram-Text model gives a class accuracy of 69.5% and an overall accuracy of 75.1% whereas the combined MFCC-Text model also gives a class accuracy of 69.5% but an overall accuracy of 76.1%, a 5.6% and an almost 7% improvement over current benchmarks respectively. The proposed models can be used for emotion-related applications such as conversational chatbots, social robots, etc. where identifying emotion and sentiment hidden in speech may play a role in the better conversation [5]. Automatic Speech Emotion Recognition helps us to build more pleasant and natural conversation partners. One of the most useful method is to use a single HMM for each emotion to be represented. The features

typically used for the classification of emotions include: - Spectral Rolloff, MFCC Features, Pitch and Magnitudes [7].

3 SPEECH EMOTION RECOGNITION

Emotion recognition systems have been in existence for a long time. The ability for a machine to identify a human's emotion have always been of keen interest. Most of these emotion recognition systems made use of an image capturing device or video recorder to identify pattern differences in the user's facial expressions. However, the inconvenience caused by having to face a camera all the time, rendered such systems to become uncomfortable for many users. In this modern era where digital assistants are becoming increasingly common, there is a need for these assistants to perform better by understanding the user's mental state. The convenient medium of sound or voice is much better and natural for human beings to interact with one another. Talking can always express your absolute thoughts to the listener in the expected manner.

3.1 Dataset

Using the proper dataset that contains quality speech data is crucial in accurately portraying the emotions of the speaker. We used the RAVDESS audio dataset, which was reduced to four emotions (neutral, happy, sad, angry) for training the SER model.

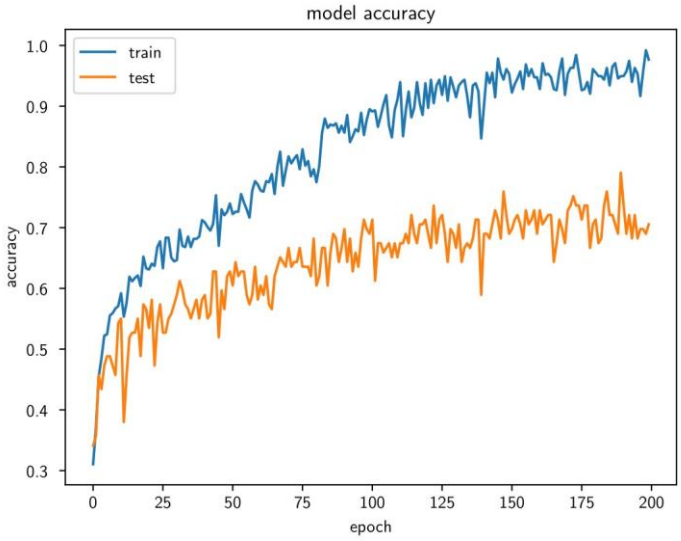
3.2 Feature Extraction

The speech signal contains many attributes that contains data that, when processed, can describe the emotion embedded in it. Selecting the features that are most relevant is one of the crucial steps to obtain maximum emotion recognition accuracy. Feature extraction is performed by reducing the data because it converts the raw speech signal into meaningful data in the form of a reduced cet of attributes while maintaining the spectral and temporal characteristics of the speech signal. We have used MFCC (Mel Frequency Cepstral Coefficients), Mel Spectrogram, Chroma Short Term Fourier Transform, Spectral Contrast, and Tonal Centroid Features as the feature extraction methods to obtain the highest emotion characteristics from the input speech audio.

3.3 Classification Model for Emotion Recognition

The emotion classification model plays the most important role in our project. The model must be capable of identifying proper attributes from the extracted features. A 5-layer Perceptron Model with 700 hidden units gave us 68.2% validation accuracy, when it was trained for 200 epochs. We also chose a 20% dropout rate to prevent over-fitting the dataset. The model could recognize the accurate emotion in most cases, but one of the primary causes for mis-classification was the quality differences of the microphone we used and the one used to create the dataset.

	angry	happy	neutral	sad
angry	27	2	2	1
happy	3	32	1	3
neutral	0	2	19	5
sad	1	5	3	23



4 CONVERSATIONAL INTELLIGENCE

The Conversational Intelligence used here is an LSTM (Long Short-Term Memory) model that classifies the intent of the user's message and generates appropriate responses while recognizing special situations to invoke custom Python and JavaScript program of associated components. This Conversational Intelligence System is the brains behind the personality of the journal. This system works in real-time, interfaced with a web page via Django Framework.

The system initially performs some pre-processing tasks on the user's message. Initially, the input text is broken down into words and reverted to its dictionary form. Next, the system creates a vector of words from the messages. The system is now ready to analyze the contextual meaning of the sentence and generate responses. Classification of the intent is performed on the input message with an error threshold of 0.25. The system's pre-trained machine learning model is used for the classification purposes. The intent generated from this step is used to generate responses. This system is programmed with some additional capabilities to analyze the intent and choose when to use the emotion classification data to generate responses. It also has capabilities to navigate through different features of the journal.

5 DJANGO FRAMEWORK

Django is a python-based web framework. It is free and open-source making it highly popular and actively backed by developer communities. It is robust and reliable making it the perfect tool to integrate a python backend to a simple web interface. Python powers most of the highly advanced artificial intelligence and machine learning systems and Django provide a window to provide these facilities via a simple web application. Django is a server-side framework. It helps the user to reduce the overhead of running complex machine learning systems to the server and lets the user enjoy the results.

In this project Django is used to deploy a frontend web application. This is used to record the user's message, convert it to text and serve it to the python backend. Using Django helped us to serve the program as a lightweight web application while the complex machine learning programs are running at the server. It reduces the load on the client system without compromising any functionalities. It also makes the system platform independent and any device with an html5 browser and JavaScript support can be used to access the application.

6 Diary Management and Sentiment Tracking System

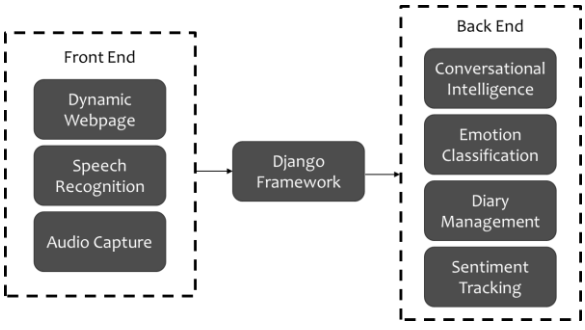
The diary management and sentiment tracking system is a collection of python programs that runs in the backend to manage the user's diary and track the emotional statistics. It works along with the speech emotion recognition (SER) system to generate positive responses to boost the emotional characteristics of the person. The system records the diary as a file and generates a report on the emotional characteristics of that person on that particular day.

The sentiment tracking system safely stores the emotional statistics in a database and uses this database to generate report. The database used here is postgreSQL. The diary management system is only invoked when the conversational intelligence navigates to the diary mode upon recognizing an appropriate message. The sentiment tracking always runs in the background to track any emotions detected during the conversation.

7 Proposed System

The intelligent personal journal that we designed and prototyped, works with a python backend and a web application-based frontend deployed via Django framework. This approach is used to execute a program with machine learning capabilities on a client system without requiring a lot of computational power.

The workflow starts from the user interacting with the dynamic web application that is programmed with HTML, CSS and JavaScript. The user will be greeted with a clock interface and two initiation buttons. When the user starts the initiation of the program the speech recognition and audio capture modules starts execution. These modules are written in JavaScript. This will continuously render the user's audio input and generate corresponding text output wherein the audio capture module will be recording the whole conversation. Once the end button is pressed the conversation is transmitted to the server side programs via the Django framework. The Django framework allows us to create the backend operations on a python program. In the server side the data is initially received by the views program. This is a default python program for Django Framework. From here the audio data is passed on to the emotion classification system. The rendered text data is passed on to the conversational intelligence system to generate the intent. Meanwhile the emotion classification system classifies the audio into any of the 4 programmed classes. These classes are happy, sad, angry and neutral. The sentiment tracking system will transfer these data to the database to generate the report. The conversational intelligence will analyze the text and determine whether to use the emotion classifier data to generate the response. If so, the system will choose a suitable response according to emotional characteristics and contextual meaning of the user's conversation. The system also determines when to go into different modes according to the user's intent. The system goes into diary mode when the user asks it to do so or when the user is about to talk about his day. In diary mode the entire conversation is safely recorded to a file without performing any operations on the data considering its sensitivity. These data are managed by the diary management system to create a date named file that records the diary entry in traditional diary format. Any response generated in this mode will be purely based on the user's emotion and is aimed at boosting the emotional status of the person. The conversational intelligence will switch back when it detects the diary entry is over. The sentiment tracking system can also generate a status report about a person's emotional state along with a detailed bar graph.



8 Results

The system proved highly effective in understanding the intent of the user and in generating appropriate response. It also succeeded in running multiple machine learning applications on the server side to do the complex tasks while interacting with a lightweight web application with an aesthetic UI. The system proved effective in determining when to use emotion data by analyzing the intent of the user. The system maintained the diary entries in correct format and generated emotional characteristic reports when needed. The accuracy of classifying emotions was above average and the system has possibility of false positives. It was also observed that the hardware used to record the user's audio input played a crucial role in the accuracy of the speech recognition and emotion classification system. It was also noted that if a more natural emotion database could be used to train the emotion classifier model the system could be highly effective in classification.

9 Conclusion

The system proposed here is a digital twin of a personal journal with intelligent capabilities. The system could converse effectively with the user by determining the user's emotion

and could improve the emotional state of the user. The system is highly scalable and can be used to induce personalization to devices and services. The system could prove worthy to analyze mental health and provide assistance. It can also lead to revelations in the field of psychology. With minimal changes, this system can be embedded into any device with a microphone, as the only physical variable required is audio data. Upon further research and data collection, the system can be trained to attain a respectable accuracy in determining emotions. The system can also be altered according to the targeted audience and location.

ACKNOWLEDGMENT

The authors wish to thank our project head Ms. Jesna Mohan, our Head of Department Ms. Tessy Mathew, our principal Mr. T M George who inspired and guided us through the entire project.

REFERENCES

[1] Rajisha T. M., Sunija A. P., Riyas K. S, "Performance Analysis of Malayalam Language Speech Emotion Recognition System using ANN/SVM," *International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015)*

[2] Nithya Roopa S., Prabhakaran M, Betty.P, "Speech Emotion Recognition using Deep Learning," *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878, Volume-7 Issue-4S, November 2018

[3] Shikha Gupta, Mr. Amit Pathak, Mr. Achal Saraf, "A Study on Speech Recognition System: A Literature Review," *International Journal of Science, Engineering and Technology Research (IJSETR)*, Volume 3, Issue 8, August 2014

[4] Leila Kerkeni, Youssef Serrestou, Mohamed Mbarki, Kosai Raoof, Mohamed Ali Mahjoub and Catherine Cleder, "Automatic Speech Emotion Recognition Using Machine Learning," *IntechOpen*

[5] Suraj Tripathi, Abhay Kumar, Abhiram Ramesh, Chirag Singh, Promod Yenigalla, "Deep Learning Based Emotion Recognition System Using Speech Features and Transcriptions," Samsung R&D Institute India – Bangalore

[6] Md. Afzal Hossan, Sheeraz Memon, Mark A Gregory, "A Novel Approach for MFCC Feature Extraction,"

[7] Yafeng Niu, Dongsheng Zou, Yadong Niu, Zhongshi He, Hua Tan, "A Breakthrough in Speech Emotion Recognition Using Deep Retinal Convolution Neural Networks," 2017

[8] Anusha Balakrishnan, Alisha Rege, "Reading Emotions from Speech Using Deep Neural Network," *arXiv:1701.08071v2 [cs.CL]* 5 Jul 2018