EMOTION DETECTION TO UPGRADE CUSTOMER EXPERIENCE

An Interim Project Report

Submitted by

ABISHECK KATHIRVEL [cb.en.u4cse18404]
E V K PRANEETH [cb.en.u4cse18419]
ROHITH RAJESH [cb.en.u4cse18450]
SANJITH RAGUL V [cb.en.u4cse18453]

Under the guidance of

Mr.Sabarish B.A
(Asst. Prof, Department of Computer Sciene & Engineering)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENE & ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

Amrita Nagar PO, Coimbatore - 641 112, Tamilnadu

Review #2 – October 2021

ABSTRACT

Growing technological development business platforms have shifted to web and cloud-based environments. Every customer needs privacy with respect to the purchases and services they seek. All business centers provide customer service support to clarify their queries in services or products. The present feedback system uses QA which ends in a biased feedback that leads to inaccurate customer experience. Interaction between the customer and executive gets affected by various reasons including inappropriate executive, extended wait time, etc. This project provides a framework that identifies the mood/emotion of the customer based on the initial chat query in a chat application and then uses sentiment analysis and routes the request to the appropriate technical expert to solve the issue which in turn improves the customer experience. After the call is established, using speech recognition the system identifies the emotion of the customer for the clarification/service provided by the person and grades them automatically. This framework for improving customer experience through sentiment analysis and emotion detection involves two phases: emotion-based call routing and auto-grading of service professionals.

The project proposes a Convolutional Neural Network(CNN) based architecture for Speech Emotion Recognition and classifies speech into angry, neutral, disappointed or happy. The outcome of this proposed system is to upgrade the customer experience by analyzing the calls at the customer care center. This system can also be used in various fields other than feedback systems like in the diagnosis of physiological disorders and counseling as emotion is an important topic in psychology and neuroscience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

LIST OF SYMBOLS

Chapter 1

INTRODUCTION

1.1 Problem Definition

To provide an improved customer service experience by emotion-based call routing using Sentiment Analysis in a chat and automatic grading for the customer service executives using Speech Emotion Recognition.

1.1.1 Justification for the proposed problem

- The current customer feedback system is not accurate and the customer service executive has no idea of the mood of the customer.
- Today's KPIs use single questions at the end of customer interaction which mostly is biased by the outcome. Hence, we cannot identify the overall customer experience (the ups and downs) in the call.
- In the proposed system the customer interacts with the chat application where the system carries out sentiment analysis to identify the mood of the user.
- The call is then routed to the best executive that can handle such situations based on prior grades the model assigned to the customer service executive.
- In every call received at the call centre the system uses speech emotion recognition for automatic grading of the executive.

Chapter 2

LITERATURE SURVEY

2.1 Emotion Recognition using speech recognition using Python

Rohan et al. (2020) aimed at presenting a comprehensive review on emotion recognition through speech using various python libraries and comparison of various classifiers. The critical part in this study is to identify the properties and characteristics of the speed signal. The following were the different classifiers used to detect emotions after feature extraction: Support Vector Machine (SVM), Random Forest classifier (RF), Multiple Layer Perception (MLP), Keras, Gaussian Naive Byes classifier (GNB), K-Nearest Neighbor (KNN). The paper proposes that speech signal is divided into subparts by framing where each frame is typically 20ms. The features which are extracted from the frames are then classified into modules based on the algorithms mentioned above. The variation of emotions is clearly identified using speech signal spectrum. After comprehensive study, the paper concludes that the highest accuracy of 82 was obtained upon training with the random forest classifier model.

2.1.1 Commonalities/differences

This paper has details about the models which we plan to use in our speech emotion recognition module. The novelty in our approach is the use of multi-modal methods where the accuracy we obtain would be based on the combined results as a result of analyzing audio-visual features.

2.2 Multi-Modal Emotion Recognition from Speech and Facial Expression Based on Deep Learning

Cai et al. (2020) aimed at developing a multi-modal emotion recognition model using facial expression benefits from the complementary information of audio-visual features. In this paper the pre-processing is followed by use of deep neural networks to extract features. The paper uses an emotion database recorded by university of south California. It contains about 12 hours of audiovisual data, namely video, audio and voice text, facial expressions, which is 10 actors (5 females and 5 male actors) in the lines or impromptu scenes, leading to emotional expression. This paper uses confusion matrix as the evaluation index of the algorithm model. The paper uses CNN and LSTM to learn global and context high-level speech emotion features, and design multiple small-scale kernel convolution blocks to extract facial expression features.

2.2.1 Commonalities/differences

The multi-modal emotion recognition used here is using the speech and face expression features that our project proposes to use. The novelty in our idea is in making use of the data obtained from the model to carry out emotion-based call routing. We also plan to have separate modules for audio and visual emotion recognition unlike the method followed here.

2.3 Ordinal Learning for Emotion Recognition in Customer Service Calls

Han et al. (2020) aimed at using a Consistent Rank Logits (CORAL) based model for ordinal speech emotion recognition. The VG-Gish is reformed in a such a way that the consecutive outputs are designed to deal with binary speech emotion recognition subtasks and the final ordinal result that is generated is based on the series of subtasks. This paper uses a call center dataset divided into three partitions with an 8:1:1 split (i.e., 3,655 utterances for the training set, 458 for the development set, 424 for the test set). The database is created from recorded customer support calls in Chinese

(8 kHz, mono). It consists of 129 conversations in total. Each utterance is rated on a three-point scale: 1-non-negative,2-Somewhat Negative, and 3-Obviously Negative. The paper successfully shows that the VG-Gish CORAL model improves performance as compared to the generic VG-Gish model. The main reason for this improvement is that the coral strategy imposes a greater penalty when larger classification errors are detected.

2.3.1 Commonalities/differences

This paper talks about using speech emotion recognition using models that we intend to use in our project. The novelty in our project is that we use sentiment analysis and facial expression-based emotion recognition and we are going to provide rating to the customer care executive on a scale of 5.

2.4 Speech Sentiment and Customer Satisfaction Estimation in Social Bot Conversations

Kim et al. (2020) aimed at showing the advantage of Bidirectional Long Short-Term Memory Networks (BLSTM) over static models through correlation analysis Customer Satisfaction (CSAT) and mean data. In addition, Kim Y, Levy J, Liu Y (2020) evaluated regression models to predict CSAT based on embeddings provided by automatic sentiment analysis system. Alexa Prize (AP) Social Data includes 6308 AP conversations and 93,671 utterances, corresponding to an average conversation's length of 14.85 utterances. The paper addresses: Activation (excited vs calm), Valence (positive vs negative) and satisfaction as the different sentiment dimensions. Kim Y, Levy J, Liu Y (2020) trained the acoustic and lexical using both acoustic and lexical cues at the utterance level. This paper uses static (SVM and RBF kernels) and temporal regression models to predict estimated CSAT score. Kim Y, Levy J, Liu Y (2020) proposed a method to automate generation of sentiment embeddings to construct model that predicts CSAT with very high accuracy.

2.4.1 Commonalities/differences

This paper talks about sentiment analysis and distinguishes different models and their accuracy which we intend to use in our project to decide on which approach to follow to implement our Chatbot based Sentiment Analysis module. The novelty in our idea is to create our own chatbot system in which we will be performing sentiment analysis with more sentiments other than activation and valence as mentioned in the paper.

2.5 A Literature Review on Application of Sentiment Analysis Using Machine Learning Techniques

Anvar Shathik and Krishna Prasad (2020) aimed at summarizing different machine learning techniques to identify emotion/sentiment. The paper also presents the different applications of sentiment analysis as a part of their literature survey. A number of sentiment analysis models and its application in various domains gives a clearer picture of the wide range of applications of sentimental analysis possible. The paper presents a tabulated summary of research works wherein the findings and research gaps in each paper are tabulated. This paper is actually a review and an extensive literature survey of existing applications of sentiment analysis and the different methods and machine learning models used for sentiment analysis.

2.5.1 Commonalities/differences

This paper is an effort to summarize numerous researches which have already been done in the field of sentiment analysis using machine learning. This will help us decide the approach we opt in our sentiment analysis model. The novelty in our idea is to create our own chatbot system in which we will be performing sentiment analysis.

2.6 Data Set

Dataset Name	IEMOCAP
Recorded by	University of Southern California
Dataset URL	https://sail.usc.edu/iemocap/
Area	Emotion
Data type	Audio-Visual data
Number of Instances	10 (5 male actors + 5 female actors)
Number of classes in output	4 (Excited, angry, sad, neutral)
Justification	The combination of audio and visual data in a single dataset makes it ideal to train models for emotion detection from speech and visual data in a call.
Keywords	Emotional, Multimodal, Acted, Dyadic

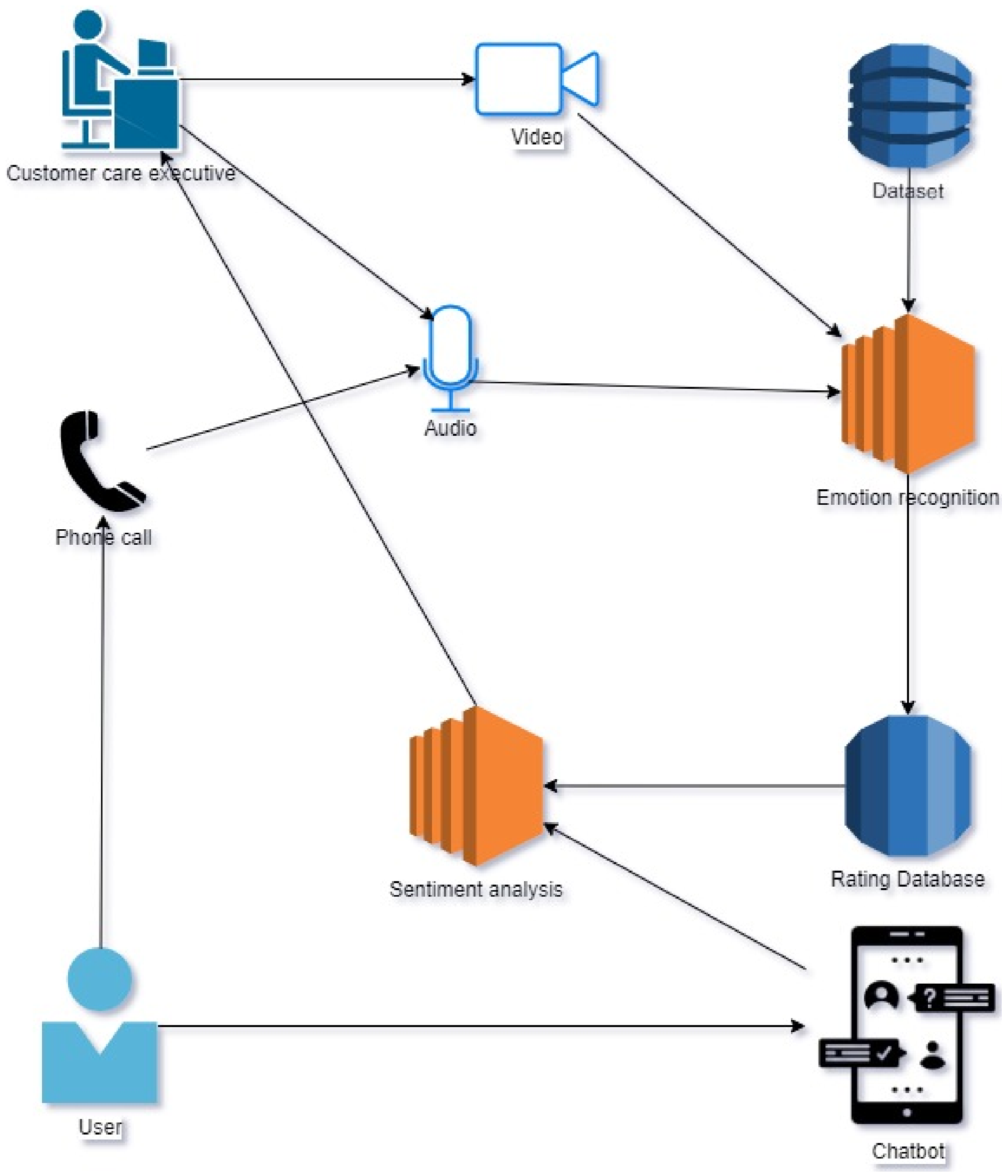
2.7 Software/Tools Requirements

- Flask (Python library) - Front End
- Librosa (Python library)- used for processing and extracting features from the audio file.
- Soundfile- Sound file reading/writing
- Pyaudio- cross-platform audio input/output stream library
- Sklearn- Predictive data analysis

Chapter 3

PROPOSED SYSTEM

3.1 Architecture Diagram



3.2 Module 1: Speech Emotion Recognition Module

In this module the goal is to develop a model which can recognize emotions of the speakers in the call at the call center. This module is prioritized as it is of high impor-

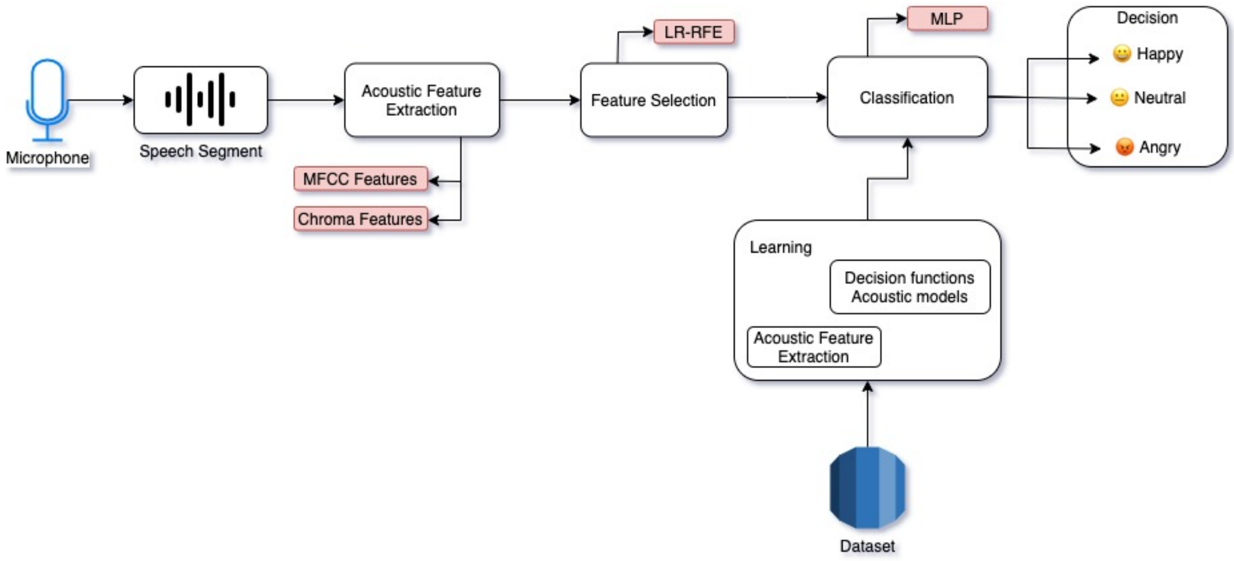


Figure 3.1: Flow diagram for speech emotion recognition

tance in the project. The results that we get from this module is used to grade executive and this will henceforth enable successful call routing.

To start with we plan to explore the existing speech emotion recognition modules and do a comparison based the results and accuracy level. The next step in this module is choosing the ideal model for our project. Once the model is chosen, we will move on to optimizing the parameters and increasing the accuracy level. The model will finally be tried out in real time wherein we will be the actors. This will be the final phase of this module.

3.2.1 Details of Submodules

- **Feature Extraction** – The speech signal contains large number of parameters that reflect emotional characteristics. MFCC and chroma(pitch) are extracted.
  - **MFCC - Mel-frequency cepstrum coefficient** is the most representation of spectral property of voice signals. 15 high order features will be extracted from the 60-dimensional MFCC feature Vector

- **Feature Selection** – Done to reduce running time of learning algorithm using Recursive feature elimination(RFE) to select the best or reject the worst performing feature. This will improve classification Accuracy. LR-RFE or SVM-RFE can be used.

- **Classification** – Multi Layer Perceptron(MLP) Classifier will be used.

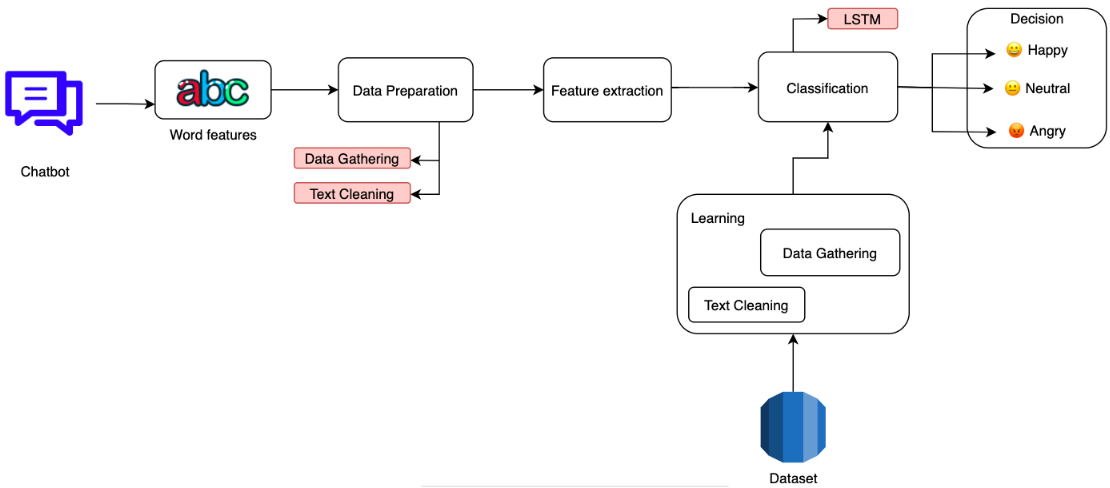


Figure 3.2: Flow diagram for sentiment analysis

3.3 Module 2: Sentiment Analysis

This module is entirely based on the data we obtain from our chat application. Initially we plan to use a popular dataset for the purpose of analyzing sentiments in chatbot applications. The chatbot application is initially planned to be a simple website where the user interacts with a bot before he gets connected to the call center executive. A model must be built which can accurately predict the emotions of the customer based on the data from the chatbot. The process of call routing will be done based on the results from this module.

3.3.1 Details of Submodules

The Sentiment Analysis module consists of three main sub modules namely Data Preparation, Build the Text Classifiers and Train the model.

- **Data Preparation – 1) Data Gathering 2) Text Cleaning**
  - **Data Gathering** – Data is required to analyze so gathering data from user reviews, social media using scraping tools, API's etc.
  - **Text Cleaning** – Removing stop words (a, and, or etc.), punctuations and whichever maybe irrelevant to the analysis.
- **Build the Text Classifier** – For sentiment analysis project, using LSTM layers with dropout mechanism to avoid overfitting.
- **Train the model** – Train the Sentiment Analysis Model on the whole dataset for 5 epochs and a validation split of 20%

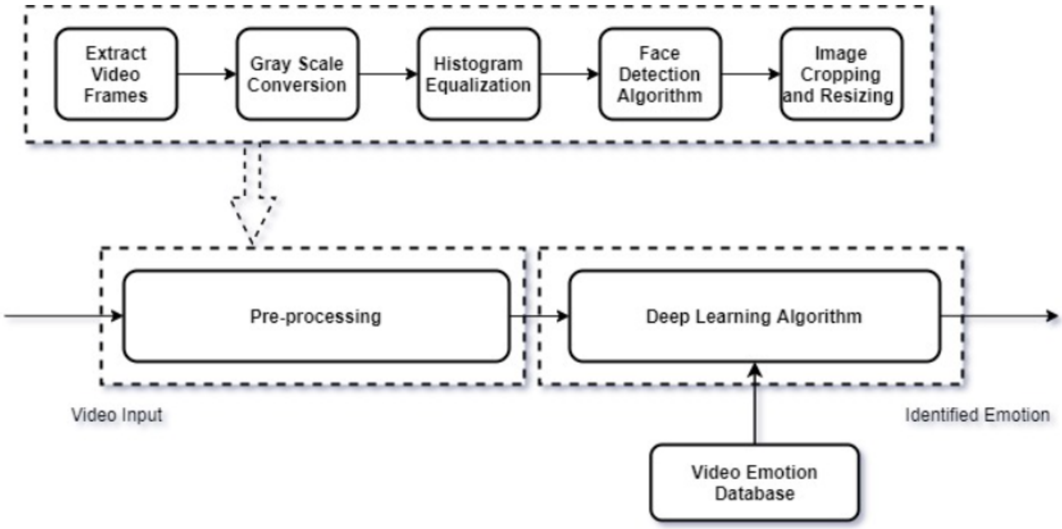


Figure 3.3: Flow diagram for salesperson video-based emotion recognition

3.4 Module 3: Salesperson video-based emotion recognition

This module brings in video-based emotion recognition wherein the emotion of the customer executive is identified by their real time facial expressions recorded by their webcam. This feature is in addition to speech emotion recognition in order to make our system more accurate.

The successful completion of this module hence ensure better results and reliability in the overall system that is built. The main tasks in this module is to explore the existing CNN models and techniques to dynamically recognize emotions and then map the model to identify emotions from a real-time video call.

3.4.1 Details of Submodules

The salesperson video-based emotion recognition module consists of three main sub modules namely Pre-Processing, Face Detection, Image cropping and resizing and Classification Model.

- **Preprocessing** – Frames are extracted from the input video which is converted to grayscale on which Histogram equalization is done to loosen up the intensity scope of the picture which reduces and computational time.
- **Face Detection** – Emotions are featured mainly from face. CNN can be used to improve the exactness of face acknowledgement calculations which would incorporate features like width, surface etc.

- **Image cropping and resizing –** The face detected is cropped to obtain a broader and clearer facial image which reduces processing times. Optimization of selected features will be done to improve accuracy.
- **Classification Model –** CNN based model to classify emotions.

3.5 Module 4: Integrating the two modules with chat application and calling system

This is the final modal wherein the task is to combine the and integrate the work done in the revious three modules. The completion of this module indicates the successful completion the solution that is proposed.

3.5.1 Details of Submodules

- **Chat and Calling System** – RASA is an opensource machine learning framework that is used to automate conversations which are trained in order to route to the correct customer care personnel according to the mood found by the sentiment analysis module.
- **Ranking System** – A Ranking system which would be updated after every call would be made which would be looked upon to do dynamic call routing according to the mood of the user.
- For the Customer Module-1 will be used and for the executive Module-3 will be used and based on the found-out emotions the Ranking system will be updated.

3.6 Conclusion

The literature survey, functional modules, the architectural diagrams of the submodules and the whole system are proposed to define a capable system that does emotion-based call routing with Speech emotion recognition and sentiment analysis

- We use sentiment analysis in the chat application in the initial phase and its results are used for emotion-based call routing to connect to the ideal person to address the situation.
- The later part involves speech emotion recognition systems wherein we analyse the call with no harm to privacy and grade the executives accordingly.
- Thereby the final goal of the project to improve overall customer experience is established.

REFERENCES

1. Anvar Shathik, J. and Krishna Prasad, K. (2020). “A literature review on application of sentiment analysis using machine learning techniques.” *International Journal of Applied Engineering and Management Letters(IJAEML).*

2. Cai, L., Dong, J., and Wei, M. (2020). “Multi-modal emotion recognition from speech and facial expression based on deep learning.” *Chinese Automation Congress (CAC).*

3. Han, W., Jiang, T., Li, Y., Schuller, B., and Ruan, H. (2020). “Ordinal learning for emotion recognition in customer service calls.” *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).*

4. Kim, Y., Levy, J., and Liu, Y. (2020). “Speech sentiment and customer satisfaction estimation in socialbot conversations.” *arXiv:2008.1237.*

5. Rohan, M., Swaroop, K., Mounika, B., Renuka, K., and Nivas, S. (2020). “Emotion recognition through speech signal using python.” *International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE).*