Volume 4 || Special Issue 10 || ICCEME 2019-2020 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

EMOTION DETECTION DURING CALL USING ARTIFICIAL INTELLIGENCE

Ashwini Madhukar Gawade[1] , Ankita Sukhdev Bobade[2] , Dipali Vitthal Dargude[3]
Monali Sampat Kharat[4] , Prof. S. A. Sirdeshpande[5]

Student, S. B. Patil College of Engineering, Indapur [1,2,3,4]

Asst. Professor , S. B. Patil College of Engineering, Indapur [5]

Abstract- The analysis of human speech is a very challenging research area as it concerns the detection of user communities. Emotions play an initial role in human interaction. The ability to understand users emotions by analyzing voice is desirable in different applications of speech recognition in emotions can be found in different areas, such as the interaction between computers and humans and call centers. Previously, emotion recognition made use of simple classifiers on bag-of-words models. However, the existing work of emotion recognition on Voice was carried out with the help of deep learning techniques on static voice data. The proposed method focuses on increasing the overall accuracy of emotion detection during calls using artificial intelligence. The overall aim is to accurately recognize the various emotions that a particular speech expresses semantically.

Keywords-*Emotion, feature extraction, Artificial intelligence, Speech to Text.*

I INTRODUCTION

Emotions can be defined as conscious affect attitudes, which constitute the display of a feeling. In recent years, a large number of studies have focused on emotion detection using opinion mining on speech. Due to some intrinsic characteristics of the voice produced during calls, such as the loudness, voice quality and casual expression, emotion recognition on them is a challenging task. Previous studies have focused mainly on lexical and deep learning methods. The performance of vocabulary-based methods largely depends on the quality of the emotional lexicon and the performance of deep learning methods depends largely on the characteristics. Therefore, we work with two classifiers that are the most famous, and have also been used before by the researchers from computational linguistics and natural language processing (NLP). Finally, Profile of Mood States is a psychological equipment that defines a four-dimensional mood state representation using text. The novel technique a Profile of Mood States generating Four-dimensional mood state representation using 65 adjectives with combination of emotions categories like, anger, happy, sad and normal. Previous work generally studied only one emotion recognition. Working with multiple classifications simultaneously not only enables performance comparisons between different emotion categorizations on the same type of data, but also allows us to develop a single model for predicting multiple classifications at the same time.

A. Problem Statement

To develop a mood state representative model for emotion recognition during calls using artificial intelligence.

B. Objectives

- To classify the user's temporary emotional state based on some input data.
- To analysis of speech as complementary to provide richer insight into the analysis of emotion.
- To detect the emotion of an on the basis of speech recognition.
- To detect four types of emotions i.e normal, happiness, anger and fear.

• Try to improve accuracy using AI techniques.

C. Motivation

The system developed based on the proposed approach would be able to automatically detect what people feel about their lives based on their voice. For example, the system can recognize:

- Percentage of people expressing higher levels of life satisfaction in one group versus another group
- Percentage of people who feel happy and cheerful
- Percentage of people who feel calm and peaceful and
- Percentage of people expressing higher levels of sad

II LITERATURE REVIEW

In this section, we briefly review the related work on Emotion Recognition system and their different techniques.

Michael Neumann, Ngoc Thang Vu: In this paper they we present our findings on how the learning of representation in a large unmarked vocal corpus can be used in a beneficial way for the recognition of the emotions of language (BE). We show that the integration of the learned representations from an unattended automatic encoder into a CNN-based emotion classifier improves recognition accuracy.

Panagiotis Tzirakis, Jiehao Zhang, Bjorn W. Schuller: In this paper, He proposed a new model for the continuous recognition of emotions from language. This model, which has been trained from one end to the other, is composed of a convolutional neural network (CNN), which extracts the characteristics of the unprocessed signal and stacks a short-term long-term 2-layer memory (LSTM), for Consider the contextual information in the data.

Po-Wei Hsiao and Chia-Ping Chen: This paper propose to integrate the attention mechanism into deep recurrent neural network models for speech emotion recognition. This is based on the intuition that it is beneficial to emphasize the expressive part of the speech signal for emotion recognition. By introducing attention mechanism, we make the system learn how to focus on the more robust or informative segments in the input signal. The proposed integration of attention mechanism on top of the baseline deep RNN model achieves 46.3% UA recall rate.

Saikat Basu, Jaybrata Chakraborty, Md. Aftabuddin: In this paperthey use 13 MFCC (Mel Frequency Cepstral Coefficient) with 13 velocity and 13 acceleration component as features and a CNN (Convolution Neural Network) and LSTM (Long Short Term Memory) based approach for classification. We chose Berlin Emotional Speech dataset (EmoDB) for classification purpose. We have approximately 80 percent of accuracy on test data.

K.Tarunika , R.B Pradeeba , P.Aruna: The primary thought of the paper is to apply Deep Neural Network (DNN) and knearest neighbor (k-NN) in arrangement of feeling from discourse particularly startling perspective. The region of use of the framework is basically worried over the human services units. The establishment of this examination has its primary firm applications in palliative consideration. Under most exact result the alarm sign are made through cloud. Numerous crude information are gathered under extraordinary accentuation methods. The voice sign are changed over to wave structure, expression level element extraction feeling acknowledgment, and ready sign creation through cloud is the arrangement of steps to be pursued.

Surekha Reddy B, T. Kishore Kumar: In this paper, a Speech Emotion Recognition (SER) system is proposed using the feature combination of Teager Energy Operator (TEO) and Linear Prediction Coefficient (LPC) features as T-LPC feature extraction. The stressed speech signals which were not accurately recognized in the previous SER systems were recognized using the proposed methods. Gaussian Mixture Model (GMM) classifier is utilized to arrange the feelings of EMO-DB database in this examination. The Stressed Speech Emotion Recognition (SSER) proposed utilizing the T-LPC include extraction procedure gained better execution contrasted with the current Pitch, LPC, and LPC + Pitch highlight based acknowledgment frameworks. This proposed emotion recognition system can be used to motivate the students by finding their emotional state providing better accuracy compared to the existing ones.

Li Zheng, Qiao Li, Hua Ban , Shuhua Liu: In this paper, a new network model (CNN-RF) based on A neural convolution network combined with a random forest is proposed. First, the neuronal convolution network is used as an extractor of features to extract the characteristics of vocal emotions from the normalized spectrogram, a random forest classification algorithm is used to classify the characteristics of vocal emotions. The result of experiment shows that CNN-RF model is superior to the traditional CNN model. Secondly, Improved the Record Sound command box of Nao and applied the CNN-RF model to Nao robot.

O. Irsoy and C. Cardie : The paper builds up a perform various tasks DNN for learning assignments over different undertakings, not just utilizing tremendous measures of cross-task information, yet in addition profiting by a regularization impact that prompts increasingly broad portrayals to help errands in new spaces. A perform various tasks profound neural system for portrayal learning, specifically concentrating on semantic grouping (inquiry arrangement) and semantic data recovery (positioning for web search) errands. Exhibit solid outcomes on question grouping and web search. Advantages are: The MT-DNN strongly performs using strong baselines across all web search and query classification tasks. Multitask DNN model successfully combines tasks as disparate as classification and ranking. Disadvantages are: The query classification incorporated either as classification or ranking tasks not comprehensive exploration work.

S. M. Mohammad and S. Kiritchenko: In article, show that feeling word hashtags are great manual names of feelings in tweets. Proposes a strategy to produce an enormous dictionary of word feeling relationship from this feeling marked tweet corpus. This is the main dictionary with genuine esteemed word feeling affiliation scores. Points of interest are: Using hashtagged tweets can gather a lot of marked information for any feeling that is utilized as a hashtag by tweeters. The hashtag emotion lexicon is performed significantly better than those that used the manually created WordNet affect lexicon. Automatically detects personality from text. Disadvantages are: This paper works only on given text not synonym of that text.

B. Nejat, G. Carenini, and R. Ng : The paper focuses on studying two fundamental NLP tasks, Discourse Parsing and Sentiment Analysis. The improvement of 3 independent recursive neural nets: for the key sub-obligations of discourse parsing, specifically structure prediction and relation prediction; the 1/3 internet for sentiment prediction. Advantages are: The latent Discourse features can help boost the performance of a neural sentiment analyzer. Pre-training and the individual models are an order of magnitude faster than the Multi-tasking model. Disadvantages are: Difficult predictions to multi-sentential text.

III PROPOSED APPROACH

The framework of speech emotion Detection is shown in Figure 1. Text recognize from human speech using raspberry pi through feature extraction techniques. Human Mood States is a psychological instrument for

Volume 4 || Special Issue 10 || ICCEME 2019-2020 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

assessing the individual's mood state. It characterizes 65 descriptive words that are evaluated by the subject on the five-point scale. Each adjective contributes to one of the four categories. The higher the score for the adjective, the more it contributes to the overall score for its category, except for relaxed and efficient whose contributions to their respective categories are negative. Mood states combines these ratings into a four-dimensional mood state representation consisting of categories: anger, happy, sad and normal. Comparing to the original structure, we discarded the adjective blue, since it only rarely corresponds to an emotion.

A. System Architecture:

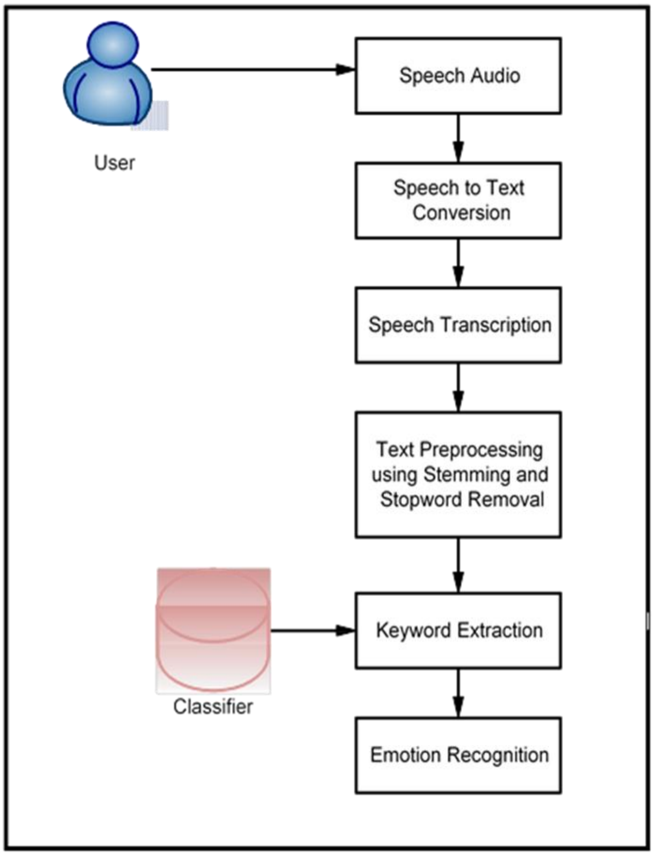


Figure 1. System Architecture

B. Algorithm

Multi-Nomial Naïve Bayes Classifier Algorithm

Function Train Naïve Bayes(D,C)

Returns $log\ P(c)\ and \log P(w|c)$

Steps:

1. For each class $c \in C$
2. Calculate $P(c)$ terms
3. N_{doc}= number of documents in D
4. N_c= number of documents from D in class c
5. $logprior[c] \leftarrow log \frac{N_c}{N_{doc}}$
6. $v \leftarrow vocabulary\ of\ D$
7. $bigdoc[c] \leftarrow append(d)$
8. For $d \in D$ with class C
9. For each word w in V
10. Calculate $P(w|c)$ terms
11. $count(w|c) \leftarrow \#\ of\ occurrences\ of\ w\ in\ bigdoc[c]$
12. $loglikelihood[w,c] \leftarrow log \frac{count(w,c)+1}{\sum_{w' in V}(count(w',c)+1}$
13. $return\ logprior, loglikelihood, V$

IV RESULTS AND DISCUSSION

The experimental result evaluation, we have notation as follows:

TP: True positive (correctly predicted number of instance)

FP: False positive (incorrectly predicted number of instance),

TN: True negative (correctly predicted the number of instances as not required)

FN false negative (incorrectly predicted the number of instances as not required),

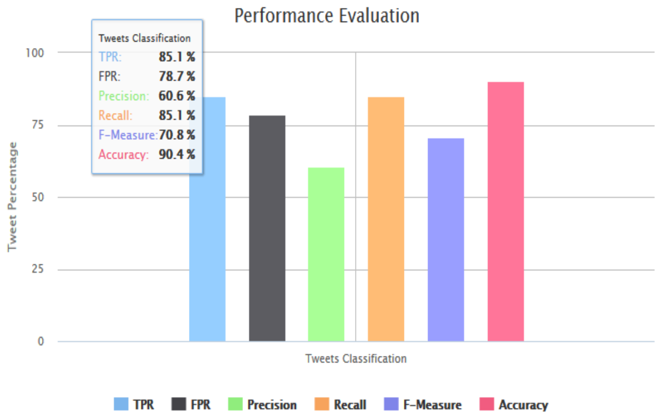
On the basis of this parameter, we can calculate four measurements

Accuracy = TP+TN÷TP+FP+TN+FN

Precision = TP ÷TP+FP

Recall= TP÷TP+FN

F1-Measure = 2×Precision×Recall ÷Precision+ Recall.



Graph 1: Performance Evaluation

Parameters	Percentage
TPR	85.1
FPR	78.7
Precision	60.6
Recall	85.1
F-Measure	78.8
Accuracy	94.4

V CONCLUSION

The technique of detecting the mood or condition of a person through voice is an emerging thought where the usefulness of this process is inevitable, and will share its uses with many sectors from medical to information technologies. This project implements an Emotion Recognition on Speech using novel technique a Profile of Mood States (POMS)using multinomial naïve Bayes represents four-dimensional mood state representation using 65 adjectives with combination of emotions categories like happy, sad, anger and normal. These POMS classifies the emotions with the help of bag-of-words.

REFERENCES

[1] Michael Neumann, Ngoc Thang Vu,” IMPROVING SPEECH EMOTION RECOGNITION WITH UNSUPERV”IEEE 2019.

[2] Panagiotis Tzirakis, Jiehao Zhang, Bjorn W. Schuller “END-TO-END SPEECH EMOTION RECOGNITION USING DEEP NEURAL NETWORKS”IEEE 2018

[3] Po-Wei Hsiao and Chia-Ping Chen” EFFECTIVE ATTENTION MECHANISM IN DYNAMIC MODELS FOR SPEECH EMOTION RECOGNITION”IEEE 2018

[4] Saikat Basu, Jaybrata Chakraborty, Md. Aftabuddin” Emotion Recognition from Speech using Convolutional Neural Network with Recurrent Neural Network Architecture” International Conference on Communication and Electronics Systems (ICCES 2017)

[5] K.Tarunika , R.B Pradeeba , P.Aruna” Applying Machine Learning Techniques for Speech Emotion Recognition” ICCCNT 2018.

[6] Surekha Reddy B, T. Kishore Kumar” Emotion Recognition of Stressed Speech using Teager Energy and Linear Prediction Features” 2018 IEEE 18th International Conference on Advanced Learning Technologies

[7] Li Zheng, Qiao Li2, Hua Ban , Shuhua Liu1” Speech Emotion Recognition Based on Convolution Neural Network combined with Random Forest”IEEE 2018

[8] O. Irsoy and C. Cardie, “Opinion Mining with Deep Recurrent Neural Networks,” in Proc. of the Conf. on Empirical Methods in Natural Language Processing. ACL, 2014, pp. 720–728.

[9] S. M. Mohammad and S. Kiritchenko, “Using Hashtags to Capture Fine Emotion Categories from Tweets,” Computational Intelligence, vol. 31, no. 2, pp. 301–326, 2015.

[10] B. Nejat, G. Carenini, and R. Ng, “Exploring Joint Neural Model for Sentence Level Discourse Parsing and Sentiment Analysis,” Proc. of the SIGDIAL 2017 Conf., no. August, pp. 289–298, 2017.