JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

RECOGNITION OF HUMAN EMOTIONS FROM VOICE SAMPLES BY ML TECHNIQUES

Akshay Mohan[1], Pranav B S[1], Hridya S B[1], Subahana N[1], Jithin Jacob[2]
[1]UG Scholar, Department of Computer Science and Engineering,
[2]Asst. Prof, Department of Computer Science and Engineering,
Dr. APJ Abdul Kalam Technological University, Kerala, India

Abstract: Even though creating effective classifiers has mostly relied on models that incorporate audio data, speech emotion recognition may be difficult. In the human lifespan, emotions play a significant part in communication, and in the modern digital world of remote communication, the detection and analysis of an equivalent are crucial. Since emotions are subjective, detecting them may be difficult. Concerning how to gauge or classify them, there is no widespread agreement. A speech emotion recognition system is described as a collection of techniques that classify and process speech signals to identify any emotions that may be present. By examining the acoustic characteristics of the audio data of recorded speech, we determined in this study to identify underlying emotions in Emotion is a natural and hereditary component of human conduct. The overall method of communication. Humans are educated to recognize a wide range of emotions via experience, which helps us to be more logical and perceptive. However, in the case of a machine, it can readily comprehend content-based information such as that presented in text, audio, or video but is still unable to grasp the depth of the content. Three types of features make up a speech: lexical features (the vocabulary used), visual features (the speaker's facial expressions), and acoustic features (sound properties like pitch, tone, and jittered.)

I. INTRODUCTION

Since spoken and auditory components make up emotional discourse, our model employs dual recurrent neural networks (RNNs) to encode the knowledge from audio and text sequences to combine this knowledge and predict the emotion class. This architecture uses the knowledge included in the information more extensively than models that focus just on audio aspects since it analyses speech data from the amplitude to the language level. To investigate the effectiveness and characteristics of the suggested model, numerous experiments are carried out. Our suggested model can perform better than earlier state-of-the-art approaches in categorizing data into at least one of four emotion categories (i.e., happy, sad, furious, and neutral). Each linguistic unit (word, phrase, or speech) is associated with one emotion from a predetermined list of emotions, with the exact start of each such unit being recognized in the continuous auditory signal. Humans have the remarkable power to modify discussions in a way that uplifts both the speaker and the listener.

II. METHODOLOGY

Speech Emotion Recognition using Time Distributed 2D-Convolution layers for CAPSULENETS In this study, speech signals are used to categorize emotions using a Time Distributed 2D-Convolution layer-based Capsule Net- work (Capsule Nets). Capsule Nets are specifically made to capture spatial signals in data, however, they fall short when it comes to considering temporal cues in time series data, such as speech. Before the Capsule Nets, Time distributed 2D- convolution neural layers are included to collect both spatial and temporal cues. The Interactive Emotional Dyadic Motion Capture (MOCAP) and Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) speech data sets are utilized to test the proposed network architecture [1]. Predicting Anxiety, Depression, and Stress in Modern Life using Machine Learning Algorithms for machine learning were used to predict anxiety, depression, and stress. Data were gathered from employed and unemployed people across many cultures and communities using the Depression, Anxiety, and Stress Scale questionnaire to apply these algorithms (DASS 21). Five distinct machine learning algorithms were used to predict the occurrence of anxiety, sadness, and stress on five different severity levels. Because these algorithms are extremely accurate, they are well suited to forecasting psychological issues. Classes were determined to be imbalanced in the confusion matrix after using the various approaches. To help choose the Random Forest classifier as the highest accuracy model among the five applied algorithms, the f1 score metric was included. Automatic Assessment of Depression from Speech via a Hierarchical Attention Transfer Network and Attention Autoencoders a deep learning method for measuring the in- tensity of speech-based sadness that incorporates unsupervised learning, knowledge transfer, and hierarchical attention. Our cutting-edge method called a Hierarchical Attention Transfer Network (HATN), employs hierarchical attention autoencoders to learn attention from a source task, then transfers this knowledge into a depression analysis system after speech recognition. The usefulness of the model is demonstrated in experiments using the depression sub-challenge dataset from the 2017 Audio/Visual Emotion Challenge (AVEC) [2]. A Convenient and Low-Cost Model of Depression Screening and Early Warning Based on Voice Using Data to Improve Public Mental Health The authoritative public data collection that had passed the ethical evaluation was used as the speech data for modelling in this work. Python programming was used to extract the voice features from the voice data, which were then saved as CSV files. A large database with 1479 speech feature samples was created through data processing and used for modeling. The decision tree screening model for depression was further developed using algorithm selection and 10-fold cross-validation. On the voice data set, the experiment's prediction accuracy was 83.4%. The model's prediction results indicate that the patients can receive an early warning and timely intervention to realize personal depression health management [3]. Its ability to foster organic communication between humans and machines, automatic speech emotion recognition has gained popularity. Speech is one way to detect emotions. The speech does, however, also include some quiet that might not be indicative of passion. Eliminating silence and/or paying closer attention to the spoken segment while disregarding the silence are two approaches to increase performance. In this article, we suggest combining the silence-removal technique with an attention model to enhance speech emotion recognition ability. According to the findings, using a combination of silence removal and an attention model performs better than using either noise reduction alone or an attention model alone [4]. To comprehend expressive human communication, speech and gesture analysis must be combined because emotions are expressed through both verbal and non-verbal channels. The Speech Analysis and Interpretation Laboratory (SAIL) at the University of Southern California has compiled a new corpus called the "interactive emotional dyadic motion capture database" (IEMOCAP) to aid in these experiments (USC)[5]. Detection of major depressive disorder using vocal acoustic analysis and machine learning—an exploratory study that job 33 people (11 men) over the age of 18 were chosen, 22 of whom had previously been diagnosed with MDD and 11 healthy controls. Their speech was captured in naturalistic contexts, throughout a typical medical exam for individuals with mental illnesses, and in various settings for health controls. Third-party voices were eliminated. A vocal feature extraction approach and other machine learning classification methods were applied to the recordings. The findings demonstrated that 100 random tree models had the best categorization performances. In this exploratory investigation, the application of machine learning classifiers with vocal acoustic data seems to be very promising for the identification of major depressive disorder, but additional research with a bigger sample size will be required to corroborate our findings[6].Mental Health Detection from Speech Signal: A Convolution Neural Networks Approach Convolution neural networks (CNNs), a machine learning technique for detecting mental health disorders changing with an emotional speech, were used in this effort to develop the model. Short-Time Fourier Transform (STFT) was used in this experiment to represent the segmented speech as a spectrogram in the frequency-time domain, and these images served as input to the CNN model. It draws attention to the benefits that CNNs can provide for mental health screening. Results show that it was a worthwhile effort, and this technique can be immediately applied by contrasting it with emotive speech[7].Real-time Acoustic based Depression Detection using Machine Learning Techniques to train a classification model that determines if a person is depressed or not, auditory features are used. For training the classifiers, the DIAC-WOZ database offered with the AVEC2016 challenge is taken into account. Using the COVAREP toolkit, prosodic, spectral, and voice control features are retrieved and merged. To overcome the class im- balance, SMOTE analysis is utilized, and the SVM algorithm produces a depression classification model with 93% accuracy (DCM). A DCM and PHQ-8 questionnaire-based android app called Cured Deployed on Cloud is being developed to help people self-assess their depression. Under the supervision of a licensed psychiatrist, the program is tested using real- time data from 50 individuals, and an accuracy of 90% is attained [8].Automated Mental Illness Analysis Using Voice Samples This study attempts to determine the best method for applying artificial intelligence to calculate depression purpose of this study was to determine whether voice may serve as a for both minor and serious depression. Based on present depression status as a dimension, 93 subjects were divided into three groups: the not depressed group (n = 33), the mild depressive episode group (n = 26), and the severe depressive episode group (n = 34). This study proved that people with minor and major depression could be accurately differentiated by machine learning and further revealed vocal changes in depressive episodes. The first study on speech changes in minor depression reveals that little depression may be detectable through voice, despite the study's limitations due to the small sample size [9]. Estimating depressive status from algorithm to determine a person's depression level from their voice signal. Patients with serious depression provided voice samples for the trial. In addition, surveys about the patients' depression were gathered. The vocalizations of three different types of long vowels by the subjects were recorded as voice signals. Then, based on the speech, acoustic features were determined. The severity of depression as determined by the HAM-D score was then estimated using an algorithm created from the recorded voice samples. The outcomes showed that the algorithm did a good job of calculating the HAM-D score severity using the long vowel's acoustic characteristics. As a result, the algorithm did a good job of evaluating the gloomy mood, indicating its usefulness for estimating depression symptoms based on speech [10].

III. CONCLUSIONS

In general, the goal of this type of project would be to develop a system that can accurately recognize and classify emotions based on vocal characteristics. This could be done using a variety of machine learning techniques, such as support vector machines, decision trees, or deep learning models. One potential application of this type of system could be in the field of mental health, where it could be used to help identify and understand the emotional states of individuals who may be struggling with mental health issues. It could also be used in customer service settings, where it could be used to help identify and respond to the emotions of customers in a more personalized and effective manner. Ultimately, the success of this type of project will depend on the quality of the data used to train the model, the appropriateness of the machine learning techniques chosen, and the effectiveness of the system in accurately recognizing and classifying emotions based on vocal characteristics.

IV. REFERENCES

[1] Bhanusree Yalamanchili1 & Koteswara Rao Anne2 & Srinivas Kumar Speech Emotion Recognition using Time Distributed 2D-Convolution layers for CAPSULENETS.

[2] Anagnostopoulos CN, Iliou T, Giannoukos I (2015) Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. Artif Intell Rev 43(2):155–177 Table 15 Train and Test accuracies train Accuracy of Model with Capsule Nets with time distribution layer on IEMOCAP test Accuracy of Model with Capsule Nets with time distribution layer on EMODB 98% 88.5% Table 16 Confusion matrix Anger Happy Sad Neutral Anger 88 60 6 Happy 6 80 4 10 Sad 0 0 90 8Neutral 4 0 0 92 Multimedia Tools and Applications.

[3] Atmaja BT, Akagi M (2019) Speech Emotion Recognition Based on Speech Segment Using LSTM with Attention Model. In: Proceedings - 2019 IEEE International Conference on Signals and Systems, ICSigSys2019, pp 40–44.

[4] Busso C et al (2008) IEMOCAP: Interactive emotional dyadic motion capture database. Lang Resour Eval42:335–359.

[5] Chen M, He X, Yang J, Zhang H (2018) 3-D convolutional recurrent neural networks with attention model for speech emotion Recognition. IEEE Signal Processing Letters 25(10):1440–1444.

[6] Akhilesh Chandra Bhatnagar, R. L. Sharma, Rajesh Kumar, "Analysis of Hamming Window Using Advance Pea Windowing Method." International Journal of Scientific Research Engineering &Technology, vol.1 issue 4, pp 015020, July 2012.

[7] Practical Cryptography, "Mel Frequency Cepstral Coefficients (MFCC) tutorial." Internet:http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/,[Feb.27, 2019].

[8] Naotoshi Seo, "Project: Pitch Detection." Internet: http://note.sonots.com/SciSoftware/Pitch.html, [Feb.25, 2019].

[9] Vocal Technologies, "Pitch Detection using Cepstral Method." Internet: https://www.vocal.com/perceptualfiltering/pitch detection/, [Feb.25, 2019].

[10] Sunil Ray, Analytics Vidhya, "Understanding SupportVector Machine algorithms from examples." Internet:https://www.analyticsvidhya.com/blog/2017/09/understaingsupport-vector-machine-example-code/, Sept.13, 2017 [Mar.10, 2019].