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

Emotion Recognition Using Speech Processing

Ms. Sheetal Rajwade[*1], Mrs. G. Mani[*2]

[1]MCA Student, Department of Master of Computer Applications,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.
[2]Assistant Professor, Department of Information Technology,
Vignan's Institute of Information Technology(A), Beside VSEZ,Duvvada,Vadlapudi Post,
Gajuwaka, Visakhapatnam-530049.

Abstract:

In the realm of human-machine interface applications, the investigation into emotion recognition from speech signals has been a longstanding research focus. Emotions hold a crucial role in human mental life, serving as a medium for expressing one's perspective or mental state to others. Speech Emotion Recognition (SER) involves extracting the emotional state of a speaker from their speech signal. Among the universal emotions are Neutral, Anger, Happiness, Sadness, Fear, etc., which can be identified or synthesized by intelligent systems with finite computational resources.This study focuses on extracting speech features, including Mel-frequency cepstral coefficients (MFCC), Chromogram, Mel-scaled spectrogram, along with Spectral contrast and Tonal Centroid features. Deep Neural Network (DNN) is employed for the classification of emotions in this work.

Keywords: Human-Machine Interface,Emotional State,Speech Emotion Recognition,Neutral,Happiness,Anger,Sadness.

1. INTRODUCTION

Communication methods vary, and among them, speech signals stand out as one of the fastest and most natural ways for humans to interact. This efficiency makes speech a viable means of communication between humans and machines. Humans naturally leverage all available senses to gain maximum awareness of the emotional state conveyed in a message. While emotional detection is inherent for humans, it poses a challenging task for machines. Therefore, the goal of an emotion recognition system is to utilize emotion-related knowledge to enhance human-machine communication.

In this system, the accuracy of speech emotion recognition is directly influenced by the quality of feature extraction. The feature extraction process involves taking entire emotional sentences as units and extracting various acoustic characteristics, including time construction, amplitude construction, fundamental frequency construction, and formant construction. By contrasting emotional speech with non-emotional sentences in these aspects, the system identifies the patterns of emotional signal distribution and classifies emotional speech accordingly.

Deep Neural Network (DNN) has shown remarkable success in speech and image recognition. Still, there has been limited application of DNN in speech emotion processing. This paper introduces a method to automatically extract emotional features from audio using the librosa package in Python. A 5-layer-deep DNN is trained to extract speech emotion features, incorporating features from consecutive frames to build a high-latitude characteristic. The

SoftMax classifier layer is then used to classify emotional speech, achieving a high accuracy of 73.38% in the speech emotion recognition test.

Traditional machine learning methods like k-nearest neighbors (KNN), Hidden Markov Model (HMM), Support Vector Machine (SVM), Artificial Neural Network (ANN), Gaussian Mixtures Model (GMM), etc., are employed for emotion classification. Key challenges in speech emotion recognition systems include the signal processing unit for extracting appropriate features and the classifier for recognizing emotions. The average accuracy of most classifiers in speaker-independent systems is lower than that for speaker-dependent systems. The growing trend of automatic emotion recognition in human speech contributes to enhanced interactions between humans and machines.

2. LITERATURE SURVEY

[1] Szegedy, Christian &Toshev, Alexander & Erhan, Dumitru.Deep Neural Networks for Object Detection. 1-9.

Deep Neural Networks (DNNs) have recently shown outstanding performance on image classification tasks. In this paper we go one step further and address the problem of object detection using DNNs, that is not only classifying but also precisely localizing objects of various classes. We present a simple and yet powerful formulation of object detection as a regression problem to object bounding box masks. We define a multi-scale inference procedure which is able to produce high-resolution object detections at a low cost by a few network applications. State-of-the-art performance of the approach is shown on Pascal VOC.

Summary: This journal discusses about the Deep Neural Networks theory and object detection using DNN.

[2] Benk, Sal & Elmir, Youssef &Dennai, Abdeslem. (2019). A Study on Automatic Speech Recognition. 10. 77-85. 10.6025/jitr/2019/10/3/77-85.

Speech is an easy and usable technique of communication between humans, but nowadays humans are not limited to connecting to each other but even to the different machines in our lives. The most important is the computer. So, this communication technique can be used between computers and humans. This interaction is done through interfaces, this area called Human Computer Interaction (HCI). This paper gives an overview of the main definitions of Automatic Speech Recognition (ASR) which is an important domain of artificial intelligence and which should be taken into account during any related research (Type of speech, vocabulary size... etc.). It also gives a summary of important research relevant to speech processing in the few last years, with a general idea of our proposal that could be considered as a contribution in this area of research and by giving a conclusion referring to certain enhancements that could be in the future works.

Summary: This article helps us in understanding and using the speech recognition by machines which improves Human Computer Interactions and is also useful in our project.

[3] Ashish B. Ingale & D. S. Chaudhari (2020). Speech Emotion Recognition. International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2 Issue-1, March 2020

In human machine interface application, emotion recognition from the speech signal has been research topic since many years. To identify the emotions from the speech signal, many systems have been developed. In this paper speech emotion recognition based on the previous technologies which uses different classifiers for the emotion recognition is reviewed. The classifiers are used to differentiate emotions such as anger, happiness, sadness, surprise, neutral state, etc. The database for the speech emotion recognition system is the emotional speech samples and the features extracted from

these speech samples are the energy, pitch, linear prediction cepstrum coefficient (LPCC), Mel frequency cepstrum coefficient (MFCC). The classification performance is based on extracted features. Inference about the performance and limitation of speech emotion recognition system based on the different classifiers are also discussed.

Summary: In this paper, we learn the importance and the need of a different features in any audio or speech including mfcc, mel and other features which are used in our application for the purpose of predicting the emotions based on audio.

[4] Chenchen Huang, Wei Gong, Wenlong Fu, Dongyu Feng, "A Research of Speech Emotion Recognition Based on Deep Belief Network and SVM", Mathematical Problems in Engineering

Feature extraction is a very important part in speech emotion recognition, and in allusion to feature extraction in speech emotion recognition problems, this paper proposed a new method of feature extraction, using DBNs in DNN to extract emotional features in speech signal automatically. By training 5 layers depth DBNs, to extract speech emotion feature and incorporate multiple consecutive frames to form a high dimensional feature. The features after training in DBNs were the input of nonlinear SVM classifier, and finally speech emotion recognition multiple classifier system was achieved. The speech emotion recognition rate of the system reached 86.5%, which was 7% higher than the original method.

Summary: In this paper, we learned the importance of DNN model and its implementation which we are going to use in our application as well.

3. EXISTING SYSTEM

In the currently available speech emotion recognition systems, the model used for emotion recognition uses traditional Machine learning algorithms like Support Vector Machines (SVM), K-Nearest neighbors (KNN) etc. The accuracies of these models are low. However, there are other models as well but they are generally trained using large datasets which takes a lot of time and hence are very complex models.

Limitations of the Existing System:

1. Lower accuracy.
2. High computational complexity.
3. Requires high performance hardware to use the application.

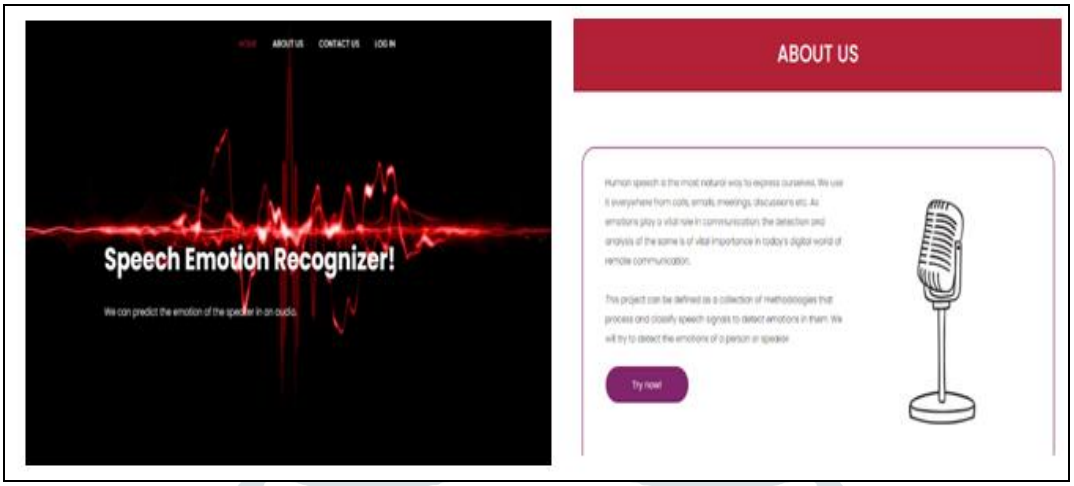
4. PROPOSED SYSTEM

We present an enhanced speech emotion recognition system utilizing deep neural networks for training. The system employs various features, including Mel-frequency cepstral coefficients (MFCC), Chromogram, Mel-scaled spectrogram, Spectral contrast, and Tonal Centroid, to extract comprehensive details from an audio file. These features are utilized to train a Deep Neural Network (DNN) model within a 5-layer deep neural network architecture. The dataset utilized for this study is the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS).Specifically, we focus on the speech portion of the dataset, which comprises 24 actors (gender-balanced) and consists of 1440 audio files. The trained model demonstrates the ability to classify speech audio into eight distinct emotions: neutral, calm, happy, sad, angry, fearful, disgust, and surprised.

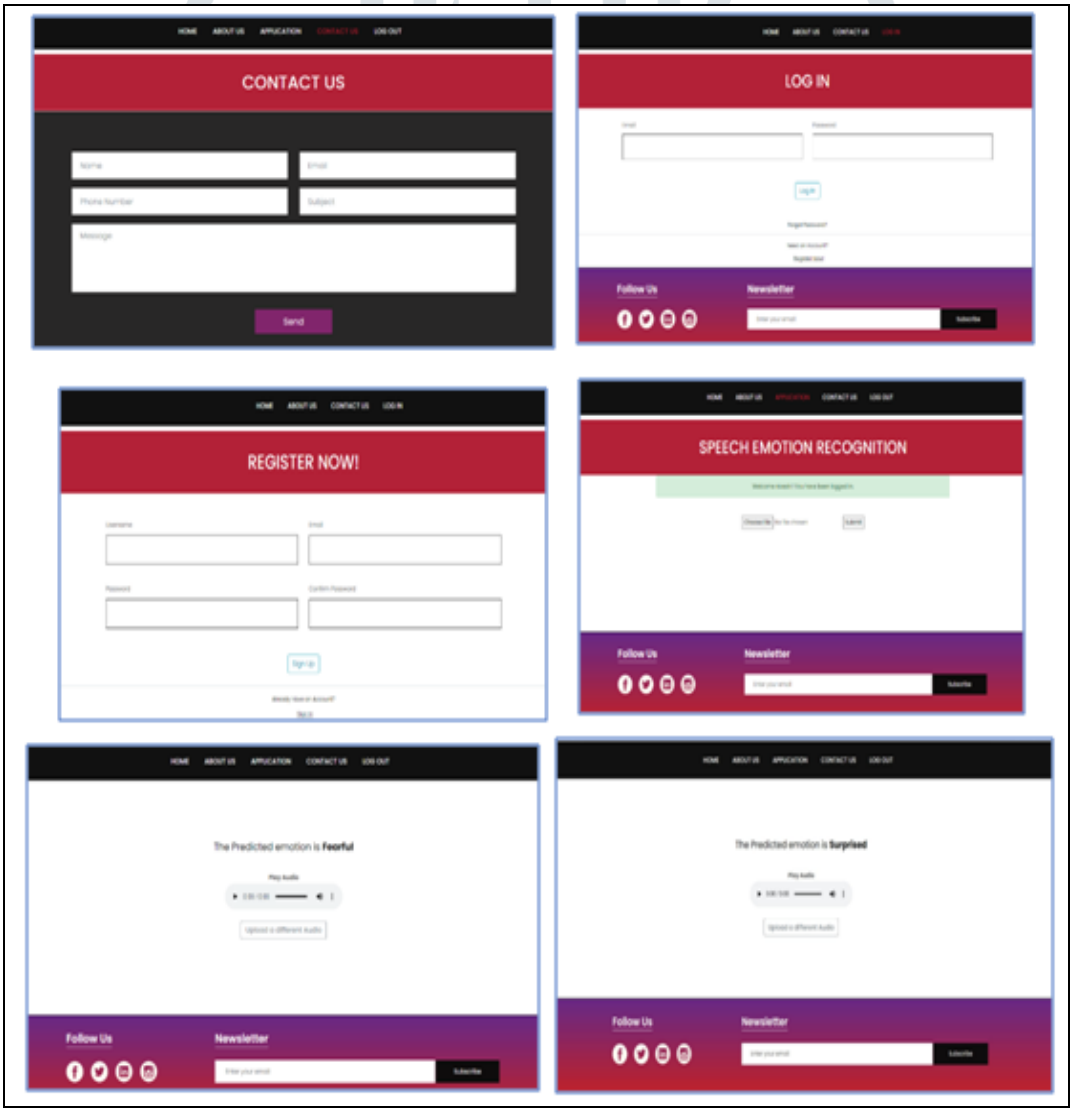
5. EXPERIMENTAL RESULTS

From the below figures it can be seen that proposed model is more accurate in order to prove our proposed system.

Home Page:



Speech Emotion Recognition:



6. CONCLUSION

In conclusion, our proposed scheme offers an effective approach for recognizing emotions from human speech, implemented through neural networks. A deep learning model was successfully developed using the deep neural network architecture to predict speaker emotions in an audio. The project was framed as a web-based application, utilizing the Flask architecture, with an incorporated user registration system in the UI. The trained model achieved a test accuracy of 73.4%.It is important to note that emotion prediction is inherently subjective, and ratings can vary among individuals for the same audio. The algorithm, trained on human-rated emotions, may yield inconsistent results. Additionally, since the model was trained on the RAVDESS dataset, variations in the speaker's accent could contribute to unpredictable outcomes, given the model's training on a North American accent database.

References

[1] Szegedy, Christian &Toshev, Alexander & Erhan, Dumitru. (2013). Deep Neural Networks for Object Detection. 1-9.

[2] Benk, Sal & Elmir, Youssef &Dennai, Abdeslem. (2019). A Study on Automatic Speech Recognition. 10. 77-85. 10.6025/jitr/2019/10/3/77-85.

[3] Ashish B. Ingale & D. S. Chaudhari (2012). Speech Emotion Recognition. International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2 Issue-1, March 2012.

[4] Chenchen Huang, Wei Gong, Wenlong Fu, Dongyu Feng, "A Research of Speech Emotion Recognition Based on Deep Belief Network and SVM", MathematicalProblems *in* Engineering, vol. 2014, ArticleID 749604, 7 pages, 2014https://doi.org/10.1155/2014/749604

[5] Livingstone SR, Russo FA (2018) The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. PLoS ONE 13(5): e0196391.

[6] M. E. Ayadi, M. S. Kamel, F. Karray, “Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases”, Pattern Recognition 44, PP.572-587, 2011.

[7] I. Chiriacescu, “Automatic Emotion Analysis Based On Speech”, M.Sc. THESIS Delft University of Technology, 2009.

[8] T. Vogt, E. Andre and J. Wagner, “Automatic Recognition of Emotions from Speech: A review of the literature and recommendations for practical realization”, LNCS 4868, PP.75-91, 2008.

[9] S. Emerich, E. Lupu, A. Apatean, “Emotions Recognitions by Speech and Facial Expressions Analysis”, 17th European Signal Processing Conference, 2009.

[10] P.Shen, Z. Changjun, X. Chen, “Automatic Speech Emotion Recognition Using Support Vector Machine”, International Conference On Electronic And Mechanical Engineering And Information Technology, 2011.