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

Speech Emotion Recognition Based on SVM Using python

R. Bhuvaneswari, S. Tamilarasi

Assistant Professor in Computer Science and Applications, D. K. M. College for women (Autonomous), Vellore

M. SC., PG Student in Computer Science, D.K.M College for women (Autonomous), Vellore

ABSTRACT

In this paper methodology for emotion recognition from speech signal is presented. Here, some of acoustic features are extracted from speech signal to analyze the characteristics and behavior of speech. The system is used to recognize the basic emotions: Anger, Happiness, Sadness and Neutral. It can serve as a basis for further designing an application for human like interaction with machines through natural language processing and improving the efficiency of emotion. In this, formant, energy, Mel Frequency Cepstral Coefficients (MFCC) has been used for feature extraction from the speech signal. Support Vector Machine (SVM) are used for recognition of emotional states. English datasets are used for analysis of emotions with SVM Kernel functions. Using this analysis the machine is trained and designed for detecting emotions in real time speech.

1. INTRODUCTION

Emotion Recognition is a recent research topic in the field of Human Computer Interaction Intelligence and mostly used to develop wide range of applications such as stress management for call centre employee, and learning & gaming software, In E-learning field, identifying students emotion timely and making appropriate treatment can enhance the quality of teaching. Main aim of HCI is to achieve a more natural interaction between machine and humans. HCI is an emerging field using which we can improve the interactions between users and computers by making computers more respond able to the user's needs. Today's HCI system has been developed to identify who is speaking or what he/she is speaking. If in the HCI system, the computers are given an ability to detect human emotions then they can know how he/she is speaking and can respond accurately and naturally like humans do. The goal of Affective computing is to recognize the emotions like Anger, Happiness, Sadness and Neutral from speech. Automatic emotion recognition and classification on voice signals can be done using different approaches like from text, voice and from human face expressions and gestures.

During present scenario, for human emotion recognition an extensive research is made by using different speech information and signal [1]. Many researchers used different classifiers for human

emotion recognition from speech such as Hidden Markov Model (HMM)[2], Neural Network (NN), Maximum likelihood bayes classifier (MLBC), Gaussian Mixture Model (GMM), Kernel deterioration and K-nearest Neighbours approach (KNN), support vector machine (SVM)[2] [3], Naive Bayes classifier.

In proposed system, basic features of speech signals like formant, Energy, and MFCC[4][5] are extracted from both offline and real time speech and they are classified into different emotional classes by using SVM classifier. Here, SVM is used since it has better classification performance than other classifiers. SVM is a supervised learning algorithm which addresses general problem of learning to discriminate between positive and negative members of given n-dimensional vectors. The SVM can be used for both classification and regression purposes. Using SVM the classification can be done linearly or nonlinearly. Here the kernel functions of SVM are used to recognize emotions with more accuracy. In human-machine interaction, The emotion recognition and classification ability is very useful. It is useful for various types of communication system such as automatic answering system, dialogue system and human like robot which can apply the emotion recognition and classification techniques so that a user feels like the system as a human.

2. LITERATURE SURVEY

Title: Comparative Analysis of Different Feature Extraction and Classifier Techniques for Speaker Identification Systems: A Review
Author: J. Kumar, Om Prakash Prabhakar

Year: 2014

Description: Speech recognition is a natural means of interaction for a human with a smart assistive environment. In order for this interaction to be effective, such a system should attain a high recognition rate even under adverse conditions. In Speech Recognition speech signals are automatically converted into the corresponding sequence of words in text. When the training and testing conditions are not similar, statistical speech recognition algorithms suffer from severe degradation in recognition accuracy. So we depend on intelligent and recognizable sounds for common communications. In this paper, we first give a brief overview of Speech Recognition and then describe some feature extraction and classifier technique. We have compared MFCC, LPC and PLP feature extraction techniques. We efficiently tested the performance of MFCC is more efficient and accurate then LPC and PLP feature extraction technique in voice recognition and thus more suitable for practical applications

Title: Speaker independent emotion recognition based on SVM/HMMS fusion system
Author: Liqin Fu, Xia Mao

Year: 2008

Description: Speech emotion recognition as a significant part has become a challenge to artificial emotion. It is particularly difficult to recognize emotion independent of the person concentrating on the speech channel. In the paper, an integrated system of hidden Markov model (HMM) and support vector machine (SVM), which combining advantages on capability to dynamic time warping of HMM and pattern recognition of SVM, has been proposed to implement speaker independent emotion classification. Firstly, all emotions are divided into two groups by SVM.

Then, HMMs are used to discriminate emotions from each group. For a more robust estimation, we also combine four HMMs classifiers into a system. The recognition result of the fusion system has been compared with the isolated HMMs using Mandarin database. Experimental results demonstrate that comparing with the method based on only HMMs, the proposed system is more effective and the average recognition rate reaches 76.1% when speaker is independent.

Title: Automatic Speech Emotion Recognition using Support Vector Machine
Author: Peipei Shen, Zhou Changjun

Year: 2011

Description: Automatic Speech Emotion Recognition (SER) is a current research topic in the field of Human Computer Interaction (HCI) with wide range of applications. The purpose of speech emotion recognition system is to automatically classify speaker's utterances into five emotional states such as disgust, boredom, sadness, neutral, and happiness. The speech samples are from Berlin emotional database and the features extracted from these utterances are energy, pitch, linear prediction cepstrum coefficients (LPCC), Mel Frequency cepstrum coefficients (MFCC), Linear Prediction coefficients and Mel cepstrum coefficients (LPCMCC). The Support Vector Machine (SVM) is used as a classifier to classify different emotional states. The system gives 66.02% classification accuracy for only using energy and pitch features, 70.7% for only using LPCMCC features, and 82.5% for using both of them.

Title: A Brief Study on Speech Emotion Recognition

Author: Akalpita Das, Laba Kr. Thakuria

Year: 2014

Description: Speech Emotion Recognition is a current research topic because of its wide range of applications and it became a challenge in the field of speech processing too. In this paper, we have carried out a brief study on Speech Emotion Analysis along with Emotion Recognition. This paper includes the study of different types of emotions, features to identify those emotions and various classifiers to classify them properly. The first part of the paper is enriched with an introductory description. Second part covers the different features along with some popular extraction method. Third part includes various classifiers used in SER and finally the conclusion part puts an end to this paper.

Title: Speaker Identification using Mel Frequency Cepstral Coefficient and BPNN
Author: Debananda Padhi, Kshamamayee Dash

Year: 2012
Description: Speech processing is emerged as one of the important application area of digital signal processing. Various fields for research in speech processing are speech recognition, speaker recognition, speech synthesis, speech coding etc. The objective of automatic speaker recognition is to extract, characterize and recognize the information about speaker identity. Feature extraction is the first step for speaker recognition. Many algorithms are suggested/developed by the researchers for feature extraction. In this work, the Mel Frequency Cepstrum Coefficient (MFCC) feature has been used for designing a text dependent speaker identification system. BPNN is used for identification of speaker after training the feature set from MFCC. Some modifications to the existing technique of MFCC for feature extraction are

also suggested to improve the speaker recognition efficiency. Information from speech recognition can be used in various ways in state-of-the-art speaker recognition systems. This includes the obvious use of recognized words to enable the use of text-dependent speaker modeling techniques when the words spoken are not given. Furthermore, it has been shown that the choice of words and phones itself can be a useful indicator of speaker identity. Also, recognizer output enables higher-level features, in particular those related to prosodic properties of speech.

3. PROPOSED SYSTEM

In proposed system, basic features of speech signals like formant, Energy, and MFCC are extracted from both offline and real time speech and they are classified into different emotional classes by using SVM classifier. Here, SVM is used since it has better classification performance than other classifiers. SVM is a supervised learning algorithm which addresses general problem of learning to discriminate between positive and negative members of given n-dimensional vectors. The SVM can be used for both classification and regression purposes. Using SVM the classification can be done linearly or nonlinearly. Here the kernel functions of SVM are used to recognize emotions with more accuracy. In human-machine interaction, the emotion recognition and classification ability is very useful. It is useful for various types of communication system such as automatic answering system, dialogue system and human like robot which can apply the emotion recognition and classification techniques so that a user feels like the system as a human.

4. MODULES

1) ACOUSTIC FEATURE EVALUATION

Speech: The primary means of communication between humans is speech. It is a complex signal which contains information about message, speaker, language, emotional states and so on.

Emotions: Emotions are defined as changes in physical and psychological feeling which influences behaviour and thought of humans. It is associated with temperament, personality, mood, motivation, energy etc.

Emotional Speech Databases: In evaluation of Emotion recognizer from speech the Main task is to check quality, naturalness and noise level of the database used in performance and efficient result estimation. When we use lower quality database for emotion recognition then there can be possibility of incorrect conclusion and result. Task of Classification also include detecting the stress of speech and it also define the type of emotion included in the database like angry, surprised, fear, happy, disgust, sad and neutral. Databases can be different types as under.

1) As Database we can consider speech samples recorded by speaking with predefined emotion from actor.

2) We can obtain Database from real life system like call centre, learning & gaming software.

3) We can also include Database with self explanatory sentiments.

In this paper English emotional speech Database in which voice samples is recorded by female and male speakers in four types of sentimental moods. Subsequently determine different audio parameters like MFCC, Formant, Energy features and stored these features vectors in database which we use for emotion recognition from speech.

Audio Feature Extraction: The speech signal contains various type of parameters from which the properties of speech are defined. Speech features generally does not very much easy to understand because of their changing behaviour and temporal adjustments make this task very tedious. In this Paper MFCC, Formant and Energy features are used. Usually the speech signal is recorded with a sample rate of 16000 Hz through microphone. The steps for calculating MFCC are described below.

2) EXTRACTION OF MEL-FREQUENCY CEPSTRUM COEFFICIENTS (MFCC)

In speech recognition the Mel Frequency Cepstral Coefficients are most widely used feature. The main purpose of using the MFCC is to mimic the behaviour of the human ears. The block diagram for MFCC is shown in Fig. 1.

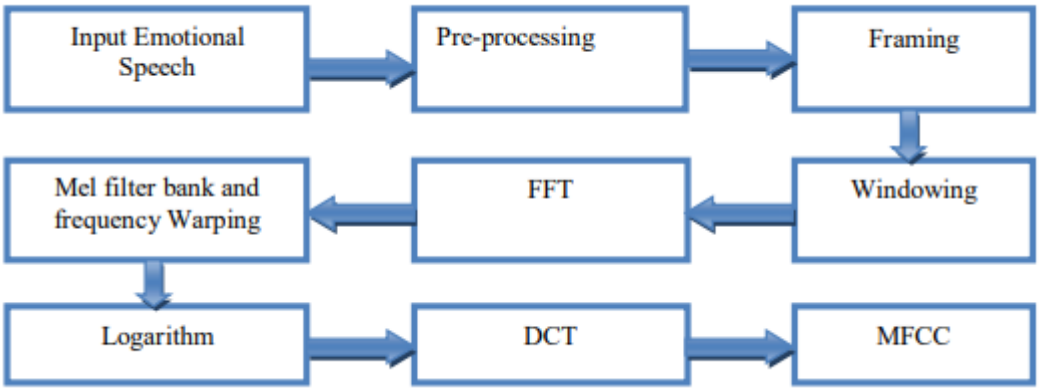

Fig-1 Block Diagram for MFCC feature extraction

Framing: In Framing, the continuous input speech is segmented into N sample per frames. The first frame consists of N samples, second frame consists of M samples after N, and third frame contains 2M and so on. Here we frame the signal with time length of 20-40ms. So the frame length of 16 KHz signal will have 0.025*16000=400samples.

Windowing: Windowing is used to window each individual frame in order to remove the discontinuities at the start and end of the frame. Hamming window is mostly used due to its relatively narrow main lobe width hence, remove distortion.

Fast Fourier Transform: FFT algorithm is used for converting the N samples from time domain to frequency domain. It is used to evaluate frequency spectrum of speech. Mel Filter Bank: In This step mapping of each frequency from frequency spectrum to Mel scale is performed. The Mel filter bank will usually consist of overlapping triangular filters with cut off frequencies which is determined by centre frequency of two filters. The Mel filters are graphically shown in Fig. 2.

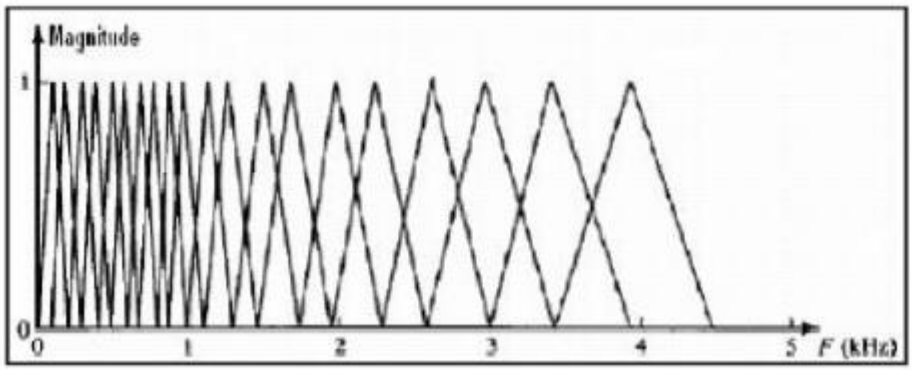

Fig-2 Mel filter bank with overlapping filters

Cepstrum: The obtained Mel spectrum is converted back to time domain with the help of DCT algorithm.

The definition of the represented of frequency in Hz to frequency in Mel scale is explained in (1) and vice versa in (2).

$$F_{Mel} = 2595 * \log_{10}(1+f_{Hz}/700) \text{ .....(1)}$$

$$f_{Hz} = 700 * (10^{F_{Mel}/2595} - 1) \text{ .....(2)}$$

5. CONCLUSION

In this paper, most recent work done in the field of Speech Emotion Recognition and Most used methods of feature extraction and several classifier performances are reviewed. In this paper we discussed about MFCC which is well known techniques used in speech recognition to describe the signal characteristics. MFCC reduce the frequency information of the speech signal into small number of coefficients which is easy and fast to compute. Success of emotion recognition is dependent on appropriate feature extraction as well as proper classifier selection from the sample emotional speech.

6. RESULT

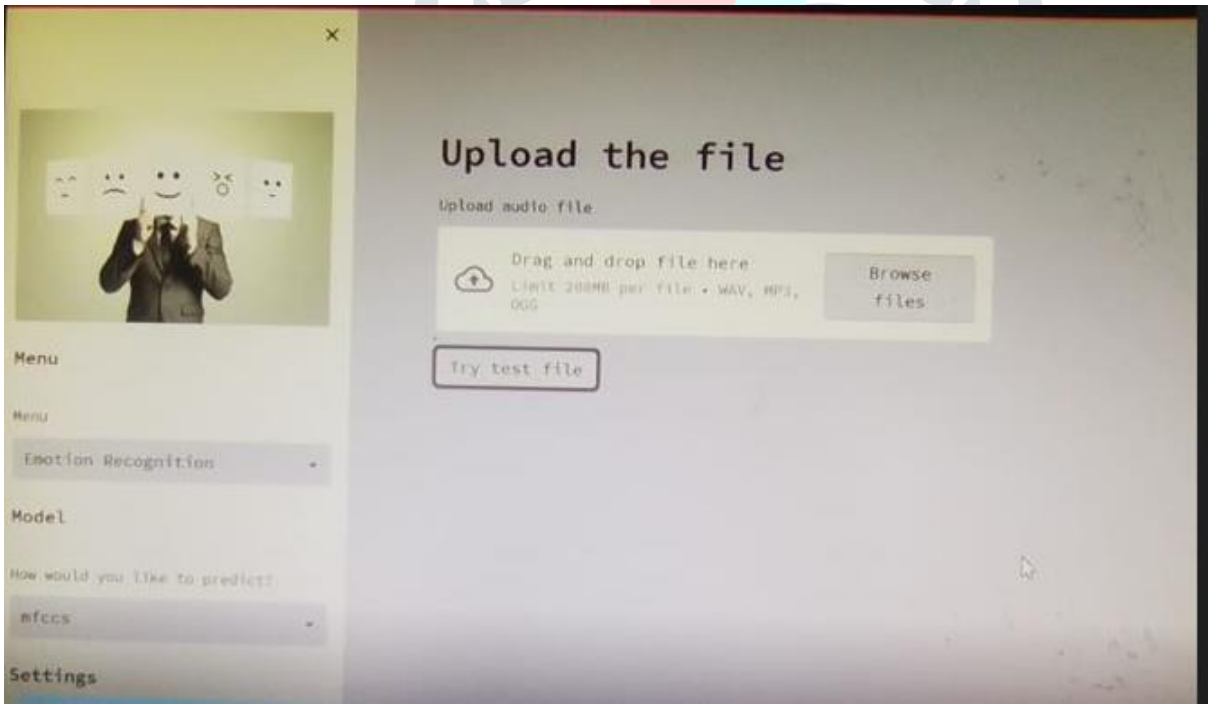

Fig 1: Upload Image

Fig 2: Analyzing Image

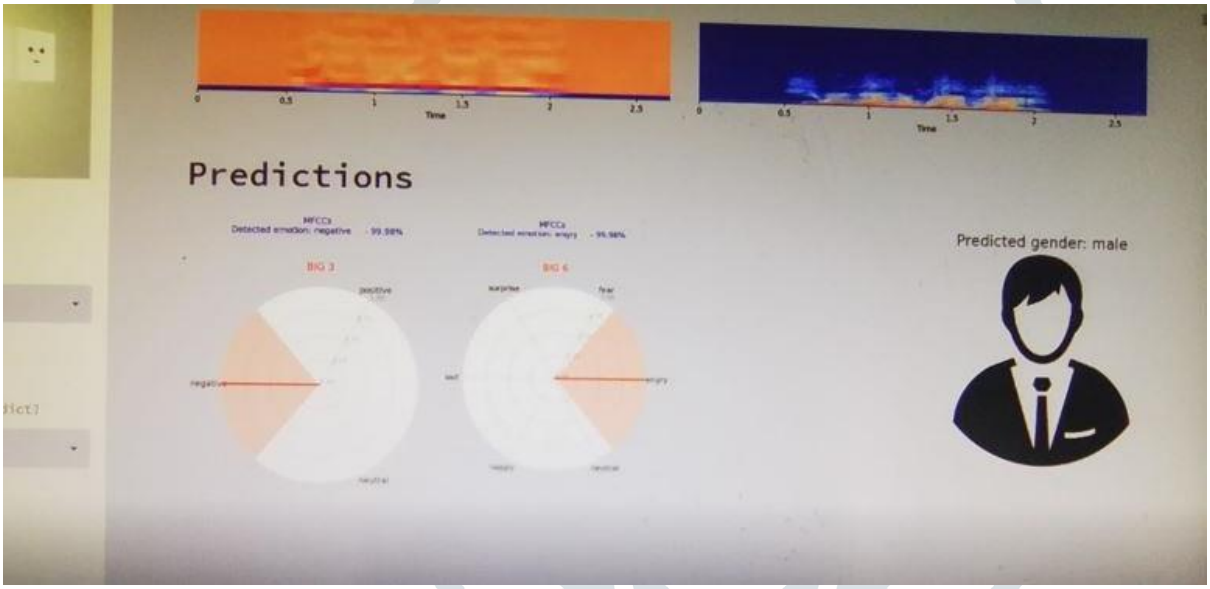

Fig 3: Predicted Image

7. FUTURE SCOPE

In Future work, It is needed to work on Emotion classification process model with SVM using different kernel functions so that it can provide better emotion recognition of real time speech and use our system in different application such as stress management for call centre employee, and learning & gaming software, In ELearning field etc. which makes our life more effective.

8. REFERENCE

1. Jeet Kumar, Om Prakash Prabhakar, Navneet Kumar Sahu, “Comparative Analysis of Different Feature Extraction and Classifier Techniques for Speaker Identification Systems: A Review”, International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE),Vol. 2, Issue 1, pg- 2760-2769, January 2014.

2. Liqin Fu, Xia Mao, Lijiang Chen “Speaker Independent Emotion Recognition Based on SVM/HMMs Fusion System” IEEE International Conference on Audio, Language and Image Processing (ICALIP), pages 61-65, 7-9 July 2008.

3. Peipei Shen, Zhou Changjun, Xiong Chen,“ Automatic Speech Emotion Recognition Using Support Vector Machine” IEEE International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT) volume2 , Page(s) : 621 - 625 , 12-14 Aug. 2011.

4. Akalpita Das, Purnendu Acharjee , Laba Kr. Thakuria , “ A brief study on speech emotion recognition” , International Journal of Scientific & Engineering Research(IJSER), Volume 5, Issue 1,pg-339-343, January-2014.

5. Kshamamayee Dash, Debananda Padhi , Bhoomika Panda, Prof. Sanghamitra Mohanty, “ Speaker Identification using Mel Frequency Cepstral Coefficient and BPNN”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 4, pg.- 326-332, April 2012.

6. Vinay, Shilpi Gupta, Anu Mehra,“Gender Specific Emotion Recognition Through Speech Signals”, IEEE International Conference on Signal Processing and Integrated Networks (SPIN), 2014 , Page(s):727 – 733, 20-21 Feb. 2014.

7. Norhaslinda Kamaruddin, Abdul wahab Rahman,Nor Sakinah Abdullah, “Speech emotion identification analysis based on different spectral feature extraction methods”, IEEE Information and Communication Technology for The Muslim World, 2014 The 5th International Conference, Pages:1-5, 2014.

8. A. D. Dileep, C. Chandra Sekhar, “GMM Based Intermediate Matching Kernel for Classification of Varying Length Patterns of Long Duration Speech Using Support Vector Machines”, IEEE Transactions on Neural Networks and Learning Systems, Volume: 25, Issue: 8,Pages: 1421 - 1432, 2014.

9. S.Lalitha, Abhishek Madhavan, Bharath Bhushan, Srinivas Saketh “Speech Emotion Recognition” IEEE International Conference on Advances in Electronics, Computers and Communications (ICAECC), Page(s): 1-4, 2014 .

10. S.Sravan Kumar, T.RangaBabu , Emotion and Gender Recognition of Speech Signals Using SVM, International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 4, Issue 3, pg.- 128-137 May 2015.