IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SPEECH EMOTION RECOGNITION USING MACHINE LEARNING

[1] Parvathy A Kumar, [2]Dr.Shine Raj G, [3] Prof Dr.Radhakrishnan B

[1]M Tech student , [2] Associate professor, [3]Head Of the Department

[1] CSE Department, BMCE,

[1] APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India

Abstract: Speech emotion recognition is extraction or identification of emotion from human speech. Feature extraction is an important part of speech emotion recognition. Machine learning is one of the best possible way to identify the emotions. This paper is a survey on the feature extraction and the classifiers used.

Keywords—Machine Learning, Feature extraction, classifiers

I. INTRODUCTION

The most fastest and natural ways of communication between the humans are speech signals. Speech emotion recognition (SER) is to recognize human emotion from speech. For human and machine interactions the fastest and most effective method is speech signal. Feature extraction is the most important part in the speech emotion recognition. For feature extraction there are many methods as prosodic features like pitch, energy, duration etc. Spectral features like MFCC, LPCC, LFPC, GFCC etc. voice quality like jitter, shimmer, HNR, normalized amplitude quotient etc. . The classifiers mostly used are classical classifiers like SVM, HMM, GMM, ANN, Decision trees etc. deep learning based enhancement techniques, classifiers based on deep learning.

II. LITERATURE STUDY

MaheshwariSelvaraj, Dr.R.Bhuvana, S.Padmaja [1], for the Speech Emotion Recognition (SER), spectral and prosodic features were used because they say that it contain more emotion information. In spectral features (Frequency domain features) have used i.e. MFCC (Mel Frequency Cepstral Coefficient) and in prosodic features they have used fundamental frequency, loudness, pitch and speech intensity and glottal parameters. For classification SVM (Support Vector Machine) and for training Radial basis function and Back propagation were used in which radial basis function produce more accurate result.

J. Nicholson, K. Takahashi and R. Nakatsu [2] the speech emotion recognition is done using one class-in-one neural networks. In this the emotions conveyed in speech was grouped into two main categories: consciously expressed emotions and unconsciously expressed emotions. In this they have selected the following eight emotional states: joy, teasing, fear, sadness, disgust, anger, surprise, neutral. The processing flow of this system was divided into two main parts: speech processing and emotion recognition. A speech input (an utterance) was inputted into the speech processing part. In this first, they calculate utterance from speech features. Next, this utterance is divided into a number of speech periods. Finally, for each speech period the speech features was extracted, and features for the utterance were compiled into a feature vector. The feature vector was then inputted into the emotion recognition part. They achieved a recognition rate of approximately 50%.

Oh-Wook Kwon, Kwokleung Chan, JiucangHao, Te-Won Lee [3] for emotion recognition, they have selected pitch, log energy, formant, mel-band energies, and mel frequency cepstral coefficients(MFCCs) as the base features, and added velocity/acceleration of pitch and MFCCs to form feature streams. Quadratic Discriminant Analysis(QDA) and support vector machine (SVM) were used to analyze the extracted features. The pitch and energy were shown to play a major role in recognizing emotion, which matches insights. In this study the accuracy is about 92.6%.

Li Zheng, Qiao Li, Hua Ban, ShuhuaLiu [4] CNN model was used as the feature extractor to extract high-order features of spectrogram. RF was used as classifier to design and implement the speech emotion recognition system. In this speech signals are divided into frames and spectrogram was calculated from emotion speech samples through framing, windowing, short-time Fourier transform (STFT) and power spectral density (PSD), and the normalized spectrogram was used as input of CNN. Speech emotion features were extracted by CNN and the output of CNN Flatten layer was input into RF classifier as eigenvectors of

speech emotion samples. In the recognition stage, the test speech signals were transformed into spectrogram and then input into the CNN-RF model classifier to recognize types of speech emotions. In this study, CNN model is used as feature extractor and combined with RF classifier. The recognition accuracy of CNN-RF is 3.25% higher than that of CNN model. So from this they suggest CNN-RF for the speech emotion recognition.

ThapaneeSeehapoch, SartraWongthanavasu [5] they have used Berlin, Japan and Thai emotion databases. In this they mainly use Speech features like Fundamental Frequency (F0), Energy, Zero Crossing Rate (ZCR), Linear Predictive Coding (LPC) and Mel Frequency Cepstral Coefficient (MFCC) from short-time wavelet signals are comprehensively investigated and Support Vector Machines (SVM) is utilized as the classification model. In this first pre-process of speech signal is done by pre-emphasis, framing and windowing and then five short time features are extracted, which are Fundamental Frequency (F0), Energy, Zero Crossing Rate(ZCR), Linear Predictive Coding(LPC) and Mel Frequency Cepstral Coefficient (MFCC). Feature normalization means that the statistical features are calculated for every window of a specified number of frames by statistical method. Feature fusion is to combine different features to build different training models. Finally, Support Vector Machines (SVM) is used as emotion classifier.

Pravina P. Ladde, Vaishali.S.Deshmukh [7] proposed the system is able to recognize four emotions (anger, happiness, sadness and neutral). This emotion recognition technique was mainly composed of two subsystems were gender recognition (GR) and emotion recognition (ER). In this they are using HMM as training algorithm and SVM as classifier. In this system the speech signals were processed, then features are extracted, then features are selected and classified. They use of serial combination of HMM and SVM classifier the accuracy of the system is about 92.50%. HMM is proved to be best training algorithm while SVM is best classification algorithm.

Leila Kerkeni, Youssef Serrestou, Mohamed Mbarki, KosaiRaoof and Mohamed Ali Mahjoub [8] have used Berlin and Spanish databases. Recurrent neural network (RNN) classifier is used to classify seven emotions found in the database for the feature extraction and classification MFCC and SVM are use. By the combination of Mel Frequency Cepstral Coefficient (MFCC), Modulation Spectral (MS) features and Recurrent Neural Network (RNN) the best result of recognition rate was about 90.05 % .

Peipei Shen, Zhou Changjun, Xiong Chen [9] proposed the speech samples are from Berlin emotional database and the features extracted from those utterances are energy, pitch, linear prediction cepstrum coefficients (LPCC), Mel Frequency cepstrum coefficients (MFCC), Linear Prediction coefficients and Mel cepstrum coefficients(LPCMCC). In this Support Vector Machine (SVM) is used as a classifier to classify different emotional states. In this it mainly consist of four modules: emotional speech input, feature extraction, SVM based classification, and recognized emotion output. They have adopted the support vector machine (SVM) to classify the speech emotion. 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.

Bjorn Schuller, Gerhard Rigoll, and Manfred Lang [10]. the system which deals with two methods. The first method global statistics framework of an utterance is classified by Gaussian mixture models using derived features of the raw pitch and energy contour of the speech signal. A second method increased temporal complexity by applying continuous hidden Markov models. In this within the first method they derived 20 features of the underlying introduced raw contours. Pitch related features, Energy related features, Processing of the derived features. The features are freed of their mean value and normalized to their standard deviation. They are classified by single state HMM's (GMM), which are able to approximate the probability distribution function of each derived feature by means of a mixture of Gaussian distributions. Each emotion is modeled by one GMM in our approach.

Tin LayNwe, Say Wei Foo, Liyanage C. De Silva [11]. They used a text independent method of emotion classification of speech is proposed. To represent the speech signals the proposed method makes use of short time log frequency power coefficients (LFPC) and the classifier used is discrete hidden Markov model (HMM). The emotions are classified into six categories. The emotions used are Anger, Disgust, Fear, Joy, Sadness and Surprise. A database consisting of 60 emotional utterances, used to train and test the proposed system. The resultant sequence of codes for each utterance is then submitted to the HMM classifier. Results show that the proposed system yields an average accuracy of 78%and the best accuracy of 96% in the classification of six emotions. This is beyond the 17% chances by a random hit for a sample set of 6 categories. From this LFPC is a better choice as feature parameters for emotion classification than others.

Chenchen Huang, Wei Gong, Wenlong Fu, and Dongyu Feng [12] they proposed a new method of feature extraction, using DBNs in DNN to extract emotional features in speech signal automatically. To extract speech emotion feature and incorporate multiple consecutive frames to form a high dimensional feature by training a 5 layers depth DBNs, then it is input to nonlinear SVM classifier, and then the speech emotion recognition multiple classifier system was achieved. They combined deep belief network and support vector machine (SVM) and proposed a classifier model which is based on deep belief networks (DBNs) and support vector machine (SVM).The speech emotion recognition rate of the system improved to 86.5%, which was 7% higher .

Table 1: Comparative study of various algorithms in literature review

Author	Year	Features	Classifiers
MaheshwariSelvaraj, Dr.R.Bhuvana, S.Padmaja	2016	MFCC	SVM
J. Nicholson, K. Takahashi and R. Nakatsu	2020	MFCC	Class-in-one neural N/W
Oh-Wook Kwon, Kwokleung Chan, JiucangHao, Te-Won Lee	2003	MFCC	QDA and SVM
Li Zheng, Qiao Li, HuaBan, ShuhuaLiu	2018	STFI, PSD	CNN
ThapaneeSeehapoch, SartraWongthanavasu	2020	LPC,ZCR, MFCC	SVM
Pravina P. Ladde, Vaishali.S.Deshmukh	2015	MFCC	SVM, HMM
Leila Kerkeni, Youssef Serrestou, Mohamed Mbarki, KosaiRaoof and Mohamed Ali Mahjoub	2018	MFCC	RNN
Peipei Shen, Zhou Changjun, Xiong Chen	2011	LPCC, LPCMCC, MFCC	SVM
Bjorn Schuller, Gerhard Rigoll, and Manfred Lang	2014	LPCC	GMM,HMM
Tin LayNwe, Say Wei Foo, Liyanage C. De Silva	2003	LFPC	HMM

III. CONCLUSION AND FUTURE WORKS

Here it discussed about different types of machine learning algorithms for prediction of Crop Prediction. Here we use various machine learning algorithms and find the best algorithm by analyzing their features. Each algorithm has given different result in different situations. For feature extraction by using MFCC the result is maximum. According to these study the classifiers that can be used are SVM . In future the system can include both age detection and accent identification from each speech signals.

REFERENCES

[1.] Human speech emotion recognition by, MaheshwariSelvaraj, Dr.R.Bhuvana, S.Padmaja International Journal of Engineering and Technology (IJET), February 2016

[2.] Emotion recognition in speech using neural network by, J. Nicholson, K. Takahashi and R. Nakatsu. Springer, December 2020

[3.] Emotion recognition by speech signals by, Oh-Wook Kwon, Kwokleung Chan, JiucangHao, Te-Won Lee. Interspeech 2003.

[4.] Speech emotion recognition based on convolution neural network combined with random forest by, Li Zheng, Qiao Li, Hua Ban, Shuhua Liu. 2018 in Chinese control and decision conference(CCDC).

[5.] Speech emotion recognition using support vector machine by, ThapaneeSeehapoch, SartraWongthanavasu. 2020

[6.] Use of multiple classifier system for gender driven speech emotion recognition by,Pravina P. Ladde, Vaishali.S.Deshmukh. 2015 International conference

[7.] Speech emotion recognition: Methods and cases studyby,Leila Kerkeni, Youssef Serrestou, Mohamed Mbarki, KosaiRaoof and Mohamed Ali Mahjoub. 10th International Conference on Agents and Artificial Intelligence , January 2018

[8.] Automatic speech emotion recognition using SVM by, Peipei Shen, Zhou Changjun, Xiong Chen.2011

[9.] Speech emotion recognition using deep neural network and extreme learning machine by, Kun Han, Dong Yu, Ivan Tashev.2014
[10.] Hidden Markov model-based speech emotion recognition by, Bjorn Schuller, Gerhard Rigoll, and Manfred Lang.2003
[11.] Speech emotion recognition using hidden markov by, Tin Lay Nwe, Say Wei Foo, Liyanage C. De Silva.
[12.] A research of speech emotion recognition based on deep belif network and SVM by, Chenchen Huang, Wei Gong, Wenlong Fu, and Dongyu Feng.
[13.] Machine learning based speech emotion recognition system by,Dr.Yogesh Kumar, Dr. Manish Mahajan.
[14.] Emotion recognition in speech signal: Experimental study, development, application by, Valery A. Petrushin