ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Implementation of Speech Emotion Recognition Based On SVM with Kernel Using MATLAB

Ritu D.Shah[1], Dr. Anil C.Suthar[2]

M.E Student, Dept. of Communication system Engineering, L.J. Institute of Engineering & technology, Ahmedabad, India [1]

Director, L.J. Institute of Engineering & technology, Ahmedabad, India[2]

ABSTRACT: In this paper we have presented methodology for emotion recognition from speech signal. In this system some of acoustic features are extracted from speech signal to analyse the characteristics and behaviour of speech. The system is used to recognize the four basic emotions: Angry, Happy, Neutral and Sad. 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 approach, formant, energy, Mel Frequency Cepstral Coefficients (MFCC) has been used for feature extraction from the speech signal. Support Vector Machine (SVM) with different kernel functions are used for recognition of emotional states. Standard EMA datasets are used for analysis of emotions with SVM Kernel functions. Result obtained from proposed approach show that emotion recognition accuracy obtained by using RBF kernel is better than the linear, polynomial, and MLP kernel functions.

KEYWORDS: Support vector Machine, Mel Frequency Cepstral Coefficients, speech signal, emotion analysis, Radial basis function, Multilayer Perceptron kernel.

I. 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 In robotics systems, stress management for call centre employee, and learning & gaming software, In E-learning field, i.e. identifying students emotion timely and making appropriate treatment can enhance the quality of teaching. The 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 Humans 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 Angry, Happy, Sad and Neutral from speech. Automatic Emotion recognition and classification in 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. Many researchers used different classifiers for human emotion recognition from speech such as Hidden Markov Model (HMM), Neural Network (NN), support vector machine (SVM), Maximum likelihood bayes classifier (MLBC), Gaussian Mixture Model (GMM), K-nearest Neighbours approach (KNN), Naive Bayes classifier[1].

In this approach, basic features of speech signals like formant, Energy, and MFCC are extracted from speech [2] 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 [3][4]. 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

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

human – machine interaction emotion recognition and classification ability is very useful. It is also 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.

II. RELATED WORK

Applications of emotion classification based on speech have already been used to facilitate interactions in our daily lives. For example, In call centres apply emotion classification to prioritize impatient customers. As another example, a warning system has been developed to detect if a driver exhibits anger or aggressive emotions. Emotion sensing has also been used in behaviour studies acoustic features have been extensively explored in both the time domain (energy, speaking rate, duration of voiced segments, zero crossing rate, etc.) and the frequency domain (pitch, formant, Mel-frequency cepstral coefficients, etc.). In our work, we only choose the most basic features: energy, formants, and MFCC. This reduces the computational complexity of the approach and can lead to both energy and bandwidth savings when the voice is captured on mobile devices. Commonly used classifiers for human emotion recognition from speech such as Hidden Markov Model (HMM), Neural Network (NN), Maximum likelihood bayes classifier (MLBC), Gaussian Mixture Model (GMM), Kernel deterioration and K-nearest Neighbours approach (KNN), support vector machine (SVM), Naive Bayes classifier. We choose SVM as our basic classifier because of its ease of training and its ability to work with any number of attributes.

In SVM, kernel functions are used to map data to a higher dimensional feature space without losing the originality. This conventional method of using kernel functions in SVM is to run simulations on training sets and find the kernel function which attains the highest averaged classification accuracy for the given problem. The most commonly used kernel function for SVM is Linear, Polynomial, radial basis function (RBF), sigmoid kernel or MLP kernel [7]. The contributions of the Speech emotion recognition are as follows: 1) To obtain the maximum efficiency using the performance of SVM kernel method for each individual technique 2) Consideration of a cut-off value in each technique so the classification having better confidence level is selected and those with lesser confidence value are discarded as 'not classified'. We have used standard EMA speech database in this approach of emotion recognition and classification [8]. The accuracy of emotion recognition can be made better by increasing the value of minimum confidence cut-off value.

III. ACOUSTIC FEATURE EVALUATION

Here, we classify the emotion of each speech utterance in the standard EMA database. Each utterance is between three or four seconds in length. We separate each utterance into 60 ms segments with 10 ms time shifts. We only choose the most basic features: energy, formant (frequency for the first four formants) and MFCC (12 MFCC values) [2][5]. Since the change in acoustic features is also related to emotional states, we include the energy difference as additional features. The 18 acoustic features employed are as follows:

Energy: energy represents the loudness of the speech. We calculate the energy for each segment by taking the summation of all the squared values of the sample's amplitudes.

Energy difference: the difference of energy values between two neighbouring segments. More fluctuations may indicate active emotions, such as happy and anger.

Formant (frequency for the first four formants, thus four features)**:** formants are determined by the shape of the vocal tract, and are influenced by different emotions. For example, high arousal results in higher mean values of the first formant frequency in all vowels, whereas positive valence results in higher mean values for the second formant frequency. We use the popular linear predictive coding method for formant calculation.

MFCC (12 MFCC values, thus 12 features)**:** MFCC coefficient of speech has a vital component in audio signals because of its simplicity in calculation, good ability to extract the feature from speech, efficient technique and also has the advantage like anti-noise etc. The actual frequency is measure in the hertz but the pitch frequency is measure on a scale which is known as the Mel frequency scale which has the frequency variations less than 0.1 KHz and logarithm variations greater than 0.1 KHz. The frequency on Mel scale can be calculated using the formula:

$$MEL(F) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \ldots\ldots\ldots\ldots\ldots (1)$$

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

The below mentioned steps in Fig.-1 are used for process of MFCC feature extraction:

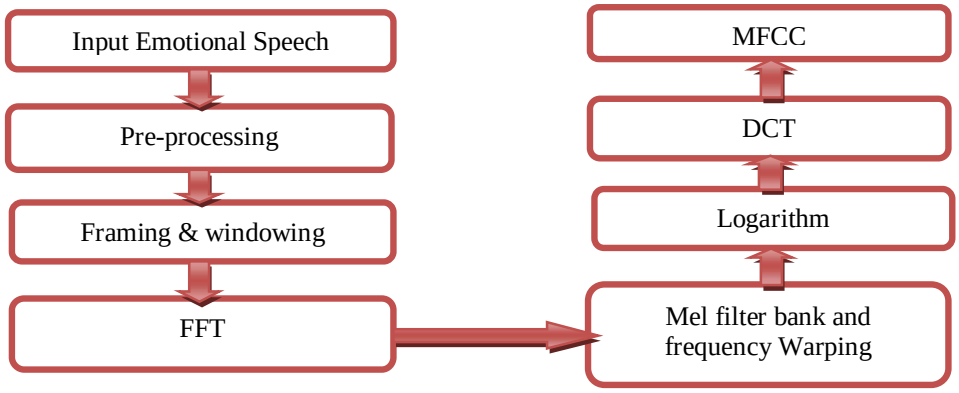

Fig. 1: MFCC feature extraction process

The reality that the human observation about the frequency information in a speech signal is not linear there is a requirement for a mapping scale. There are various scales for getting right perception about frequency content. The scale used in this experiment is the Mel scale. This scale is represented a measured frequency of a pitch to a corresponding pitch calculated on the Mel scale. The definition of the represented of frequency in Hz to frequency in Mel scale is explained in Eq.2 and vice versa in Eq.3.

$$F_{mel} = 2595 \log_{10}\left(1 + \frac{f_{Hz}}{700}\right) \ldots\ldots\ldots\ldots (2)$$

$$f_{Hz} = 700\left(10^{\frac{F_{mel}}{2595}} - 1\right) \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

We find these 18 features for each 60 ms segment of the speech sample, and then we calculate the mean, the maximum, the minimum, the range, and the standard deviation for each feature, resulting in 18 x 5 = 90 attributes that are sent to the classifier.

IV. A FRAME WORK OF SPEECH EMOTION CLASSIFICATION USING SVM CLASSIFIER

The SVM approach is a high dimensional vector supervised learning method that is based on emotion assumptions [7]. It predicts that the presence (or absence) of a specified feature of a class is not related to the presence (or absence) of all other features. It is very simple to program and execute it, its parameters are simple to assume, learning or training is very fast and effective even on very large databases and its accuracy is comparatively better in comparison to the other techniques. The emotion recognition process along with training and testing phases is shown in figure 2.

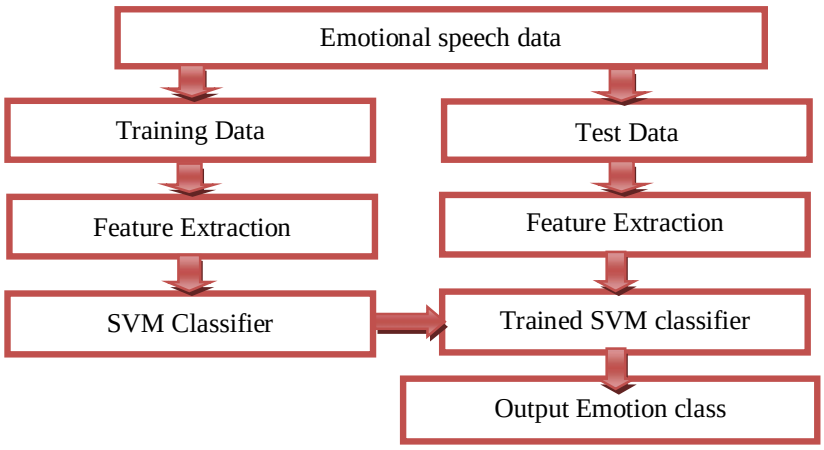

Fig. 2 Emotion classification process model

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

V. RESULTS AND EVALUATION

Selection of appropriate kernel function:

The approach is implemented with sequential minimization optimization (SMO) method for optimized parameter setting of SVM classifier on learning and preparation of dataset. First of all the audio samples are processed and selection of appropriate audio parameters for extraction is decided. The basic emotions those have to be identified are: angry, happy, neutral, and sad. The cut-off parameter for initial experiments has been setup to 0 and after that the experiment is repeated by increasing the cut-off parameter and measuring the confidence level of emotional classification. The amount of datasets is increased by applying the multi-fold cross-validation techniques for inherent combinations like: linear, polynomial, MLP and RBF.

$$\text{CL recall } X_i = \frac{Ctpi}{Ctpi+Cfni} \ldots\ldots\ldots\ldots \quad (4)$$

$$\text{CL accuracy } X_i = \frac{Ctpi+Ctni}{Ctpi+Ctni+Cfpi+Cfni} \ldots\ldots\ldots\ldots \quad (5)$$

Where, Ctni, Ctpi, Cfni and Cfpi represents the number of true negative classifier, true positive, false –ve and false +ve instances correspondingly. The recall value and accuracy level of SVM classifier is defined by above equations. Results obtained for CL accuracy (%) using different kernel functions are represented in Table 1. Here, we have used 7 fold cross validation as shown in Fig.3 in which result obtained for RBF kernel is shown. While training separate SVM classifiers, this has been assumed that there should be enough quantity of learning instances for training model otherwise it would produce inaccurate output. As an illustration while the classifier model is learning for 'sad or not' emotion, 'not sad' emotion comprising of all other emotions. According to the accuracy level and recall value appropriate kernel expression for all classifier models have been represented as per given conditions in table 1. The output waveform representation of feature extraction is shown in Fig. 4.

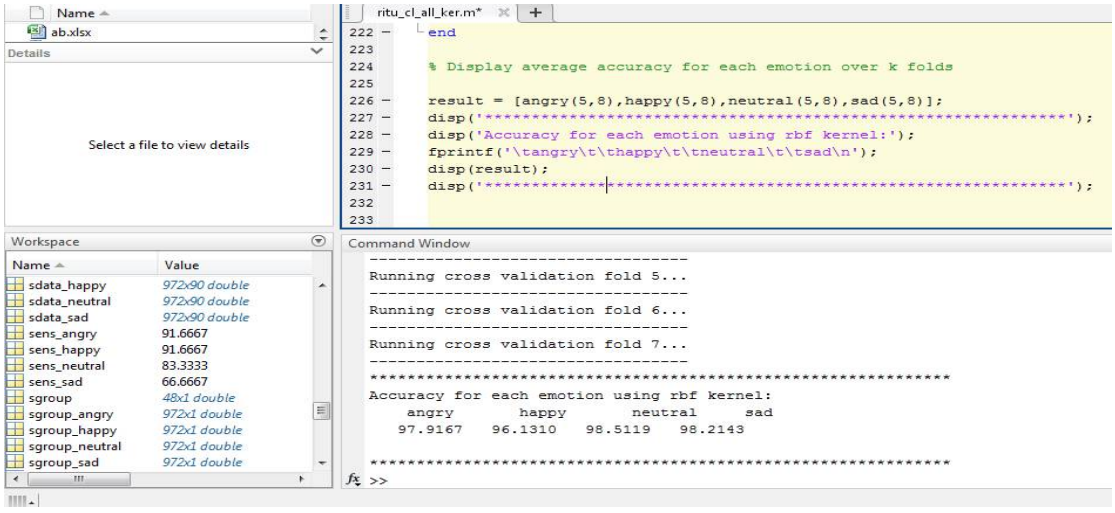

Fig-3 Accuracy obtained using RBF kernel function

Table-1 Results obtained for CL-accuracy (%) using different kernel functions

Kernels	Angry	Happy	Neutral	Sad
Linear	97.91	93.45	93.75	93.75
Polynomial	94.64	94.34	93.75	96.42
RBF	97.91	96.13	98.51	98.21
Sigmoid	30.35	46.13	51.48	47.32

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

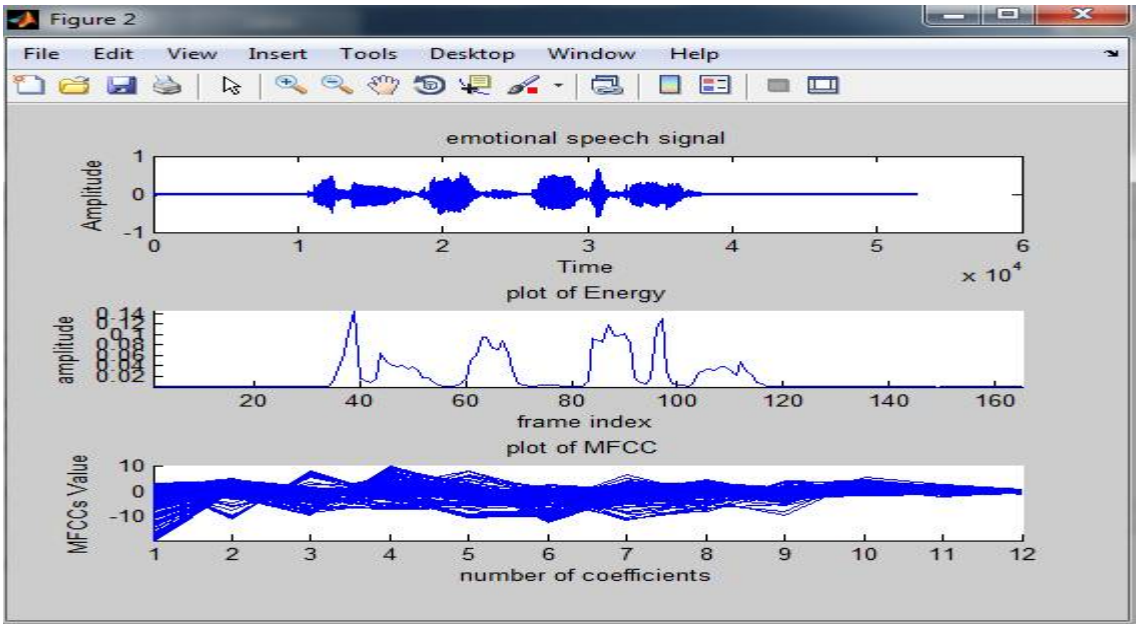

Fig.4- Result for feature extraction

VI. CONCLUSION AND FUTURE WORK

In this paper, speech emotion recognition and classification method using multi-class SVM has been presented. Speech samples from standard EMA database are used in analysing recognition and classification of emotions from audio samples. It has been observed that by enhancing the cut-off parameter and multi cross-validation the confidence level of classifier attains best possible correct classification accuracy of the classification model. Here, we have used linear, polynomial, RBF, and sigmoid kernel functions and from result we can conclude that RBF kernel gives more classification accuracy for emotion recognition compared to other three kernel functions. Future prospects of approach include developing this idea further by taking large varieties of emotional text, audio and video samples. More acoustic features can also be included for emotion recognition from speech.

REFERENCES

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. Ritu D.Shah, Dr. Anil.C.Suthar, "Speech Emotion Recognition Based on SVM Using MATLAB" International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Vol. 4, Issue 3, pg-2916-2921, March 2016.
3. 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.
4. 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.
5. 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.
6. 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.
7. Bo Yu, Hai Feng Li, Chun Ying Fang, " speech Emotion Recognition based on Optimized Support Vector Machine" Journal of Software, Vol 7, No 12(2012), 2726-2733, Dec 2012.
8. http://xios.usc.edu/emotional_ema_form.php.
9. 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.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

10. 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.
11. 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.
12. 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 .
13. 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.

BIOGRAPHY

Ritu D. Shah received her B.E. degree in Electronics & Communication Engineering from Gujarat technological University. She is currently pursuing her M.E degree in Communication System Engineering from Gujarat technological University at L. J. Institute of Engineering & Technology, Himmatnagar, India.

Dr. Anil C. Suthar is currently Director at L. J. Institute of Engineering & Technology, Ahmedabad, India. He is Ph.D.in Electronics & communication. He has completed ME in Communication systems from L. D. College of Engineering, Ahmedabad, in year 2006. He has published 12 articles in Electronics for you and EM media Pvt. limited, most popular electronics magazine in Asia. He has guided several projects at graduate and post graduate level.