ISSN:0975-9646
Reshma et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2969-2971

Speech Emotion Recognition by Gaussian Mixture Model

Reshma[1], Maninder[2], Amarbir Singh[3]

[1,2]*Department of Computer Science Engineering, Chandigarh University, Gharuan*
[3]*Department of Mechanical Engineering, Chandigarh University, Gharuan*

***Abstract*- In the field of human computer interaction automatic speech emotion recognition is a current research topic. Emotion recognition in speech is a challenging problem because it is unclear that which features are effective for speech emotion recognition. In this paper we proposed an approach in which we extract the features of energy, spectral and acoustic domains and then merging these features by Principal Component Analysis(PCA) and then we get a new hybrid feature that feed into a Gaussian mixture model with kernel approach and analysis its accuracy, precision, recall and ROC curve**.

Keywords:- Feature Extraction, GMM, SVM, Feature Selection, PCA ,LDA

I. INTRODUCTION

The latest challenge in speech processing is how to deal with emotion of the speaker. Speech is most natural way of communication between human beings. This fact motivated researchers to think about to increases man machine interaction. To recognize emotional state of the person machine requires sufficient intelligence level. Emotion is a state of mind that are associated with feelings, opinion and different kind of behaviors. Emotion detection is a process through which the real emotion can be recognized by extracting features from the emotional speech.

It is very easy to detect emotions from speaker but it is too challenging to detect emotions through machine. Speech emotion recognition can be used for call centers and E-learning programs. In call centers main objective for speech emotion recognition is to detect satisfaction and dissatisfaction of customer where as in E-Learning programs objective is to adjust presentation style of E-tutor when learner is bored, interested or frustrated. Other applications of speech emotional recognition are Lie detection, Intelligent Toys, Psychiatric diagnosis etc.[1]

Speech emotional recognition is very challenging because it is too difficult to recognize boundaries between more than one type of emotion perceived through emotional speech. Another reason is sometimes expressing emotions depends on speaker, his or her culture or environment.

Emotion can be detected by extracting features from the samples of the speech corpus. Speech corpus contains some audio recording of sentences by the human that represents different kind of basic emotion like sad, anger, happy, neutral There are two type of information that we can conceived from emotional speech.

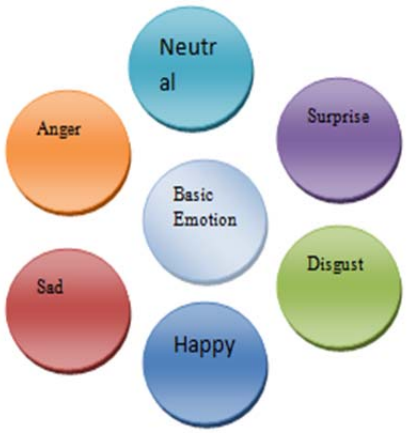

Fig 1.1 Basic Emotions

First is concerned with linguist part which refers to the acceptance of all rules of pronunciation whereas second information concludes emotional state of speaker.[3] Extracted features from emotional speech represent emotional information. Features can be categorized into two parts- spectral and prosodic features. Spectral features are energy , pitch ,intensity etc where as prosodic features are Mel frequency cepstrum coefficient , Linear prediction coefficient , Linear prediction cepstrum coefficient etc [7].

Speech emotion detection has plenty of applications in E-commerce as customer satisfaction in customer care centers , E – learning to analyze the emotional state of learner so as to improve their presenting skills, Medical Field as a psychiatric diagnosis.

II. RELATED STUDY

There are some studies in the field of speech processing that represents highly correlation between emotion detection and features of speech. Those features are related to energy , pitch and spectral measures of speech. In 2008 , Yoglin et al.[2] proposed a systematic approach of emotion detection by Audiovisual signals. Mel Frequency Cepstrum Coefficient features helps in extracting audio information while Gabor wavelet features are used to represent visual information. They compare different classifier approaches and analyze that multimodal classifier gives best accuracy.

In 2013, Tiwali et al [4] proposed a framework to improve performance of facial expression recognition. They performed experiment on eNTERFACE'05 audiovisual emotional database using six basic emotions - sad, happy,

disgust, surprise, anger, neutral. They compare multiclass classification with one to one binary classification and observe that multiclass classification yields better results.

In 2012, Yixing et al. [5] used different combinations of features like pitch , Energy , LPCC , MFCC , MEDC and found that feature combination MFCC + MEDC + Energy gives maximum accuracy on German and Chinese Database using support vector machine as a classifier.

The codebook is constructed by using feature vector model based on Mel frequency short time speech power and emotions can be recognized by Discrete Hidden Markov Model. They concluded that emotional state of speaker can be reflect significantly by information retrieved through filter bank coefficient.[6]

The affective voice content strongly affected by amplitude and frequency control of sound. So a novel approach of using amplitude and frequency-derived features is proposed and results are compared on Berlin database emotional speech and recently collected Athens emotional states inventory.[8]. Yogjing Wang et al.[9] proposed a novel approach of kernel cross multimodal factor that represent cross modal relationship between audio and visual channels respectively. An experiment of data mining on feature selection is conducted and more relevant features are selected from 1000 features derived from pitch, energy and MFCC time series by removing correlated features.[10]

III. PROPOSED APPROACH

Emotion detection is the process through which the real state of mind can be analyzed of the speaker. There are two main tasks in emotion detection. First is to train the classifier and second is to test the classifier. To train the classifier again two steps are very important. They are feature extraction and feature selection.

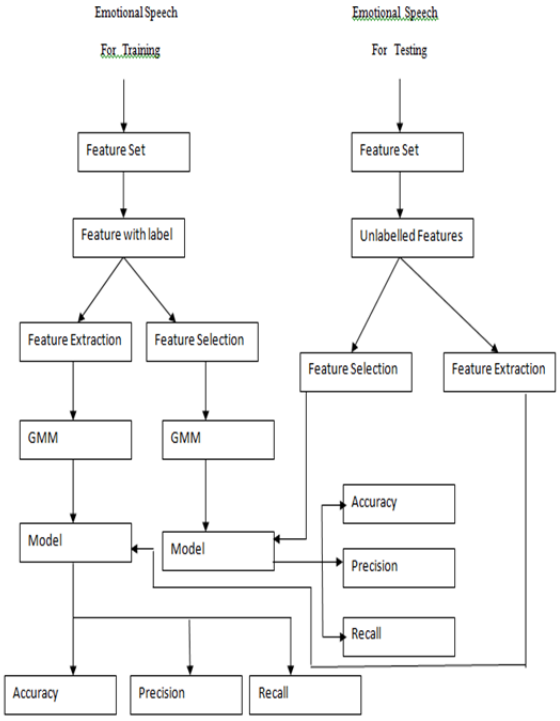

Fig 1.2 Emotion detection process

Feature extraction is the process through which different type of features can be extracted from the emotional speech that are energy entropy block, short time energy , zero crossing rate , spectral roll off , spectral centroid , spectral flux , noise to harmonic ratio , harmonic to noise ratio. We are performing feature extraction by Principal Component Analysis. PCA can reduce data and works on similarity of variances of all input speech features set. It provides a small features set from all extracted features.

Feature Selection is technique that reduce dimensionality of the Features. It helps in categorization of features. To select the features we are using Linear Discreminent Analysis. Linear disceminent analysis (LDA) performs Feature selection by reducing dimensionality of the features. LDA works on correlation coefficient of all classes and it is used to differentiate all classes. Kernel is used to check similarity between pair of inputs.

Gaussian mixture model is a classifier that is used to train the model. GMM is based on probability distribution of the features measured in a biometric system such as vocal-tract related spectral features in a speaker recognition system. It is used for representing the existence of sub-populations, which is described using the mixture distribution, within the overall population. We will compare the results with Support vector machine and Adaboost. SVM is a binary classifier to analyze the data and recognize the patterns for classification and regression analysis and Adaboost is an iterative algorithm that focuses on both continuous valued input and textual input by text categorization. It combines many simple and moderately accurate rules into a single highly accurate rule.

Algorithm 1 (PCA Algorithm for emotion detection in Speech)

Input: unlabelled speech
Output: labeled speech with particular emotion

1. For I=0 to length (speech file)
. {
Energy Features
Acoustic Features
Cepstral Features
}
2. For I=0 to length (speech features)
. {
Put in PCA
Extract Features
}
3. Featuresr(x_1 ,x_2 ,x_3,……, x_n) with labels
4. Input features in Gaussian Mixture Algorithm and train it
5 Make a Gaussian Mixture Model =X
6. For I=0 to length (speech Test)
. {
Select features
Extract by PCA
}
7. Features input in X
X given a label emotion anger, disgust , fear , happiness , neutral , sad, surprise
8. Check Precision , recall and accuracy.

Algorithm 2 (LDA Algorithm for emotion detection in Speech)

Input: unlabelled speech
Output: labeled speech with particular emotion

1. For I=0 to length (speech file)
. {
Energy Features
Acoustic Features
Cepstral Features
}
2. For I=0 to length (speech features)
. {
Put in LDA
Extract Features
}
3. Features(x_1 ,x_2 ,x_3,……, x_n) with l labels
4. Input features in Gaussian Mixture Algorithm and train it
5 Make a Gaussian Mixture Model =X
6. For I=0 to length (speech Test)
. {
Select features
Extract by LDA
}
7. Features input in X
X given a label emotion anger, disgust , fear , happiness , neutral , sad, surprise
8. Check Precision , recall and accuracy.

IV. CONCLUSION

In the field of human computer interaction automatic speech emotion recognition is a current research topic. Emotion recognition in speech is a challenging problem because it is unclear that which features are effective for speech emotion recognition. In this paper we will extract the features by PCA therefore dimension of the processing is less than before existing approaches and we will compare the results of GMM classifier with other classifiers.

REFERENCES

[1] El Ayadi, Moataz, Mohamed S. Kamel, and Fakhri Karray. "Survey on speech emotion recognition: Features, classification schemes, and databases." Pattern Recognition 44.3 (2011): 572-587.
[2] Wang, Yongjin, and Ling Guan. "Recognizing human emotional state from audiovisual signals*." Multimedia, IEEE Transactions on 10.5 (2008): 936-946.
[3] Chaspari, Theodora, Dimitrios Dimitriadis, and Petros Maragos. "Emotion classification of speech using modulation features." Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 22nd European. IEEE, 2014.
[4] Tawari, Ashish, and Mohan M. Trivedi. "Face expression recognition by cross modal data association." Multimedia, IEEE Transactions on 15.7 (2013): 1543-1552.
[5] Bozkurt, Elif, et al. "Improving automatic emotion recognition from speech signals." INTERSPEECH. 2009.
[6] Nwe, Tin Lay, Say Wei Foo, and Liyanage C. De Silva. "Speech emotion recognition using hidden Markov models." Speech communication 41.4 (2003): 603-623.
[7] Chaspari, Theodora, Dimitrios Dimitriadis, and Petros Maragos. "Emotion classification of speech using modulation features." Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 22nd European. IEEE, 2014.
[8] Yalamanchili, B. S., et al. "Non Linear Classification for Emotion Detection on Telugu Corpus." International Journal of Computer Science & Information Technologies 5.2 (2014).
[9] Wang, Yongjin, Ling Guan, and Anastasios N. Venetsanopoulos. "Kernel cross-modal factor analysis for information fusion with application to bimodal emotion recognition." Multimedia, IEEE Transactions on 14.3 (2012): 597-607.
[10] Vogt, Thurid, and Elisabeth André. "Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition." Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on. IEEE, 2005.