Performance Analysis of Different Classifiers in Speech Emotion Recognition

Ch.V.Kiranmayi[1], Dr.S.V.R.K.Rao[2], Biswaranjan Barik[3]
[1]PG.Scholar, [2]Professor, [3]Associate Professor
[123]GIET College, Rajamahendravaram

Abstract- Recognition of emotion from speech signal is a new topic in present field of research, it has plenty of applications. One of the applications is Human machine interaction (HMI). The main objective of this research work is to develop Automatic -Emotion Recognition (AER) technique to recognize emotions in Telugu. It also compared the performance of different classifiers and developed a new enhanced method for the improved performance. Spectral features will consider only few feature vectors are not sufficient for emotion classification. To extract more feature vectors the DWT is used. The effect of Cepstral coefficients in the detection of emotions is performed and also a comparative analysis of Cepstrum, Mel-frequency Cepstral Coefficients (MFCC), and pitch on emotion classification is done. The classification task is performed using artificial neural network's back propagation algorithm. The proposed method has shown better performance in identifying four emotions from Telugu database that of other classifiers (GMM, Bayesian). Classification accuracy is improved.

Keywords- AER, DWT, MFCC, Cepstrum and Pitch, GMM, Bayesian, ANN.

I. INTRODUCTION

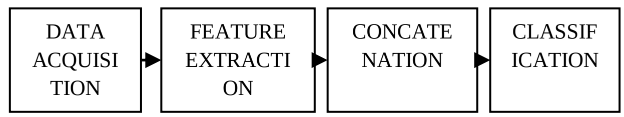
All forms of human communication carry information at two levels, the message and the underlying emotional state. Emotions are essential part of real life communication among human beings. Emotions are experienced when something unexpected happens at times. Emotion recognition is the process of examining or identifying the status of a person. The emotion recognition enable human computer or human machine interaction in future, which may possibly understand with persons or help people with communication tribulations and detection for psychiatric diagnosis, amongst other applications. Earlier research works have used different classifiers and feature extraction techniques and failed to improve accuracy and reduce misclassification. To enhance the accuracy artificial network with different nodes is deployed as classifier.

II. DATA BASE

Training and testing of the Artificial Neural Network is performed by using Telugu database. A male and a female speaker's voices of four different emotions namely neutral, angry, sad and happy are recorded. The database consists of 157 samples of both training and testing.

III. DISCUSSION

This section explains process involved in suggested method. The proposed method carried in five steps and is shown in following figure. Figure.1: Processing steps of speech emotion recognition



Data acquisition: The voice signals from different speakers are collected in this process; the data is collected so as to have four different emotions angry, sad, happy and neutral.
Feature Extraction: Spectral and prosody features are extracted using different methods. The techniques used to extract features involve DWT (Discrete Wavelet transform), Cepstrum, MFCC etc.
Concatenation: Features obtained from above process are concatenated to perform classification. The output of this process will produce a feature vector.
Classification: Classifier consider input vector and classify the emotion class based on trained data. Different classifiers used for emotion recognition

1. Gaussian Mixture Model
2. Bayesian Classifier
3. Artificial Neural network
4. Deep Networks

Gaussian Mixture model: Gaussian Mixture Models (GMMs) are among the most statistically mature methods for clustering or unsupervised learning. A Gaussian mixture model provides a good approximation of the originally observed feature probability density functions by a mixture of weighted Gaussians. Each emotion is modeled in one GMM. The decision is made for the maximum likelihood model. There are results in [11] where the authors concluded that using GMMs is a feasible technique for emotion classification.
Bayesian Classifier: Naïve Bayes classifier is based on the so-called Bayesian theorem with the naïve assumption of independence between every pair of features. This classifier in spite of the apparently oversimplified assumptions has worked quite well in many real-world situations. It is very fast and has

a good performance, better in some cases than more sophisticated methods. The Naïve Bayes model with Gaussian is equivalent to a mixture of Gaussians (GMM) with diagonal covariance matrices. The main advantages of this classifier are the conditional independence assumption, which helps to obtain a quick classification, and the probabilistic hypotheses.

Artificial Neural network: Artificial Neural Network is worked with an efficient particularly sorted out approach to manage refresh an execution point of view or to take after some clear inside basic, which is regularly proposed as the learning standard. The information/yield getting ready data are first in neural framework change, since they pass on the principal information to "locate" the perfect working point. The nonlinear thought of the neural structure managing sections (PEs) outfits the system with social occasions of versatility to achieve in every helpful sense any pined for data/yield arrange.

Deep networks: Deep Networks have been successfully used for audio analysis, speech recognition, and natural language processing [14]. Non-linear deep networks such as Boltzmann Machines [12] and Neural Networks [13] have the ability to learn a rich feature presentation in an unsupervised manner, making them very powerful. Restricted Boltzmann Machines (RBMs) form the building blocks of deep networks models. These models are trained using the Contrastive Divergence (CD) algorithm, which enables deep networks to capture the distribution so vertex features efficiently and to learn complex representations [15].

IV. RESULTS

GMM: After training and testing the emotion recognition models results are obtained in form of 4X4 matrix form. Each row of matrix represents the test data recognized by different models in form of percentage and each column of the matrix represents the trained model. In this matrix, diagonal elements show the correct classification and other elements

Emotion	Angry	Happy	Sad	neutral
Angry	53	11	10	25
Happy	0	67	22	11
Sad	6	5	52	37
Neutral	20	15	4	61

Table 1: Recognition rate of GMM

From the above table it is clear that by using GMM the system is having 53% chance to recognize the emotion type angry similarly happy 67%, sad 52%and neutral 61%.Overall accuracy rate of the system with GMM is 59%.

BAYESIAN: First of all we classify with prosodic features. Over 200 features were reduced to eight by using SFFS. The confusion matrix is shown in Table 1. The overall recognition rate with 66.7% is quite good, but mainly the discrimination between anger and happiness is bad. Happiness is least classified with a recognition rate of only 48.2%. Furthermore, the confusions between angry and anxious and between neutral with bored are noticeable. As we know from Fig. 1, these emotions do

Emotion	Angry	Happy	Sad	Neutral
Angry	67%	18%	0%	0.7%
Happy	34%	48%	0.9%	5.5%
Sad	0%	0.9%	77%	9%
Neutral	0%	6%	10%	62%

Table.2: Recognition rate of Bayesian classifier.

Table1. Classification with prosodic features only

Not differ in the activation dimension and so prosodic features cannot adequately distinguish between them. On the other hand, sad is classified best with 77.3%. In this database, sadness is spoken very slowly and also with long pauses. Hence, duration features work very well to recognize sad utterances.

DEEP NETWORKS: a hybrid model comprising of temporal generative and discriminative models for detection and recognition of emotional content in speech has shown improvement when compared with earlier classifiers. After deploying of deep network classification accuracy has been increased to 76%.

ANN: The first remarkable conclusion about these results is the fact that in all cases the low level feature seems to be useful for emotion recognition. Instantaneous features provide better performance than syllabic ones for both energy and pitch, and pitch features work also better than energy ones. As a result, the best partial combination is instantaneous pitch.

50 nodes	20 nodes	10 nodes
93%	90%	45%

Table.3: Recognition rate of ANN classifier.

Artificial neural network shown an improvement in recognition rate and it is 93%.

V. CONCLUSION

We presented a new approach to classify emotions from speech signals and successfully recognized the four emotions namely angry, sad, happy and neutral. The performance of different classifiers is compared among all classifiers artificial neural networks with back propagation algorithm shown better results. By using ANN we achieved 93% recognition rate. Future scope will be considered with different features and different classifiers.

VI. REFERENCES

[1]. S Lalitha, D Geyasruti, R Narayanan, M Shravani (2015) Emotion Detection Using MFCC and Cepstrum Features. In: Procedia Computer Science 70 (2015) 29-35.

[2]. FirozShah.A, Vimal Krishnan V.R., RajiSukumar.A, AthulyaJayakumar, BabuAnto.P “Speaker Independent Automatic Emotion Recognition from Speech:-A Comparison

of MFCCs and Discrete Wavelet Transforms" 2009 International Conference on Advances in Recent Technologies in Communication and Computing, pp:528-531,2009.
[3]. S.Kim, P.Georgiou, S.Lee, S.Narayanan "Real-time emotion detection system using speech: Multi-modal fusion of different timescale features", Proceedings of IEEE Multimedia Signal Processing Workshop, Chania, Greece, 2007.
[4]. K.V.Krishna Kishore, P. Krishna Satish "Emotion Recognition in Speech using MFCC and Wavelet Features",3rd IEEE International Advance Computing Conference.
[5]. Firoz Shah.A, Vimal Krishnan V.R, Raji Sukumar.A, Athulya Jayakumar, Babu Anto.P "Speaker Independent Automatic Emotion Recognition from Speech:-A Comparison of MFCCs and Discrete Wavelet Transforms" 2009 International Conference on Advances in Recent Technologies in Communication and Computing, pp:528 531,2009.
[6]. Batliner A, Fischer K, Huber R, et al. How to Find Trouble in Communication [J]. Speech Communication, 2003, 40,(12):pp.117~143;
[7]. Cowie R, Douglas-Cowie E, Tsapatsoulis N, et al. Emotion Recognition in Human-Computer Interaction[J].IEEESignalProcessingmagazine,Vol.18,No.1,2001 pp.32~80;
[8]. J. SirishaDevi ,Dr. Srinivas Yarramalle,Siva Prasad Nandyala "Speaker Emotion Recognition Based on Speech Features and Classification Techniques" I.J. Image, Graphics and Signal Processing, 2014, No:7, pp: 61-77 ,June 2014.
[9]. Siqing Wu, Tiago H.Falk, Wai-Yip Chan "Automatic speech emotion recognition using modulation spectral features". In: Speech communication 53 (2011) 768-785.
[10]. Michael Grimm, KristianKroschel, Emily Mower, Srikanth Narayanan"Primitives-based evaluation and estimation of emotions in speech". In: Speech communication 49 (2007) 787800.
[11]. K. E. a. K. L. D. Neiberg, Emotion Recognition in Spontaneous Speech Using GMMs, Pittsburg, 2006, pp. 809-812.
[12]. Y. Bengio, "Learning deep architectures for ai," in FTML, 2009.
[13]. L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network leaning for speech recognition and related applications: An overview," in ICASSP, 2013.
[14]. L.DengandD.Yu,"Deeplearningforsignalandinformation processing," in FTML, 2013.
[15]. G.E.Hinton, S.Osindero, andY.W.Teh, "Afastlearning algorithm for deep belief nets," in NC, 2006.