



Speech based Human Emotion Recognition Using Hybrid Classifier Technique

Shubhangi S. Jarande*, Surendra Waghmare

Dept. of E & TC, G.H. Raisoni College of Engineering and Management,
University of Pune, Maharashtra, India

Abstract: *Speech Emotion Recognition (SER) is a current research topic in the field of Human computer Interaction (HCI) with wide range of applications. The speech features such as, Mel Frequency cepstrum coefficients (MFCC), Energy is extracted from speech utterance. The Support Vector Machine (SVM), Probabilistic Neural Network is used as classifier to classify different emotional states such as anger, happiness, sadness, neutral, fear, from a database of emotional speech. The SVM and PNN is used for classification of emotions. SVM and PNN using together forms a Hybrid Classifier which increase the efficiency of the Speech Recognition System. It gives 90% classification accuracy.*

Keywords— *Speech emotion, Emotion Recognition, SVM, PNN, Energy, MFCC.*

I. INTRODUCTION

Speech Emotion Recognition is a very current research topic in the Human Computer Interaction field. As computers have turn into an integral part of our lives, the requirement has rise for a more natural communiqué interface between humans and computers. To achieve this objective, a computer would have to be capable to perceive its present situation and responds in a different way, depends on that perception. Part of this process involve understanding a user's emotional state. To build the human-computer communication more natural, it would be helpful to give computers the ability to recognize emotional situations the same way as human does. A Emotion Recognition System can be done in two ways, either by facial expressions or by speech. In the field of HCI, speech is crucial to the objective of an emotion recognition system, as are gestures and facial expressions. Speech is consider as a controlling mode to communicate with intentions and emotions. In the current years, a great deal of research has been done to recognize human emotion using speech information [12].

Many researcher explore several classification methods including the Gaussian Mixture Model (GMM), Neural Network (NN), Maximum Likelihood Bayes classifier (MLC), Hidden Markov Model (HMM), Kernel Regression, Support Vector Machine (SVM) and K-nearest Neighbors (KNN). The Support Vector Machine is binary classifier used for emotion recognition. The SVM is used for regression purpose and classification. It perform classification by constructing an N-dimensional hyper planes that optimally separate the data into categories[14]. The classification is achieved by a nonlinear or linear separating surface in the input feature space of the dataset. Its main design is to transform the original input set to a high dimensional feature space by using a kernel function, and then get optimum classification in this new feature space.

II. BASIC ARCHITECTURE

The block diagram of the emotion recognition system in the course of speech considered is shown in Fig. 1. The Emotion Recognition System Consists of five blocks such as input speech, feature extraction, feature extraction, feature selection, classifier and emotional speech at output.

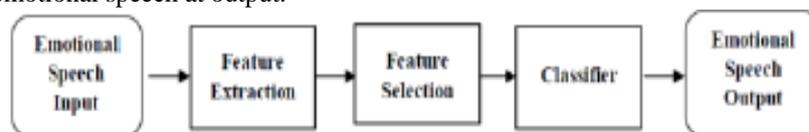


Fig. 1 Basic Block Diagram of Emotion Recognition

A. Emotional Speech as Input:

Emotional Database is very important requirement of any emotional recognition model. Database quality determines the efficiency of the system.

B. Feature Extraction and Selection:

A very important step in Human Emotion Recognition System through speech is to pick a feature which contains large amount of emotional information. In general, the features extracted using Energy, MFCC, Pitch, PLP, etc. The MFCC feature involved some steps which are as shown below:

a. Mel-frequency cepstral coefficients (MFCC)

Mel-frequency cepstral coefficients are the coefficients that collectively forms an Mel Frequency Cepstrum. They are formed from a type of cepstral depiction of the audio clip. The difference between the mel-frequency cepstrum and the cepstrum is that in the MFC, the frequency bands are placed equally on the mel scale. For better representation of sound, frequency warping can permit it. Fig. 2 shows the MFCC feature extraction process [3]. As shown in Fig. 2 feature extraction process contains following steps:

- I. Pre-processing: The continuous time signal means speech is sampled at standard sampling frequency rate. At the first stage in MFCC feature extraction is to boost up the amount of energy in the high frequencies. For pre-emphasis filter is used.
- II. Framing: In framing process speech samples are segmented by using ADC converter, into small frames which have length in between 20-40msec. Framing converts non stationary signal to quasi stationary frames.
- III. Windowing: Windowing step is used to window each of the individual frame for minimizing the discontinuities at the beginning and end of each frame.
- IV. FFT: Fast Fourier Transform algorithm is used for evaluating the frequency spectrum of speech.
- V. Mel Filter bank and Frequency wrapping: The mel filter bank having overlapping triangular filters with the cut-off frequencies determined by the centre frequencies of the two adjacent filters. The filters have fixed bandwidth on the mel scale and linearly spaced centre frequencies.
- VI. Take Logarithm: The logarithm has the result of changing multiplication into addition.
- VII. Take Discrete Cosine Transform: It is used to orthogonalise the filter energy vectors

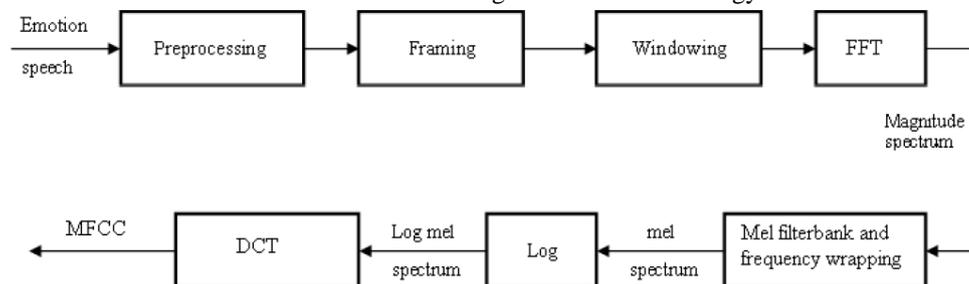


Fig. 2 Block Diagram of the MFCC Feature Extraction

b. Energy

The Energy is the most important and basic feature in speech signal. Energy oftenly referred to the intensity or volume of the speech, where it is also contains valuable information. Energy provide information that can be used to discriminate sets of emotions, but this measurement alone is not enough to discriminate basic emotions. Anger and Joy have increased energy level, where low energy level is sadness. Mean of energy is taken into consider in the human emotion recognition system

$$E_n = \sum_{n=1}^N x(n) \cdot x^*(n)$$

C. Classifier

Classifier is the most important aspect in human emotion speech recognition system. The systems performance is depends on the proper selection of classifier. There are different types of classifiers available as HMM, GMM, PNN, SVM, etc

III. PROPOSED SYSTEM

The Aim of the system design is its efficiency and simplicity. The figure 3 shows the speech recognition system. Generally the speech files are in .wav and mp3 format. In proposed system the speech files are in .wav format are used. Database is used which is created by recording the different emotional sounds. The figure 3 is as shown below which is separated in two different parts. The right hand side represent the testing part and left hand side represent the training part. System contains blocks as input speech emotion, feature extraction and training.

3. Classification

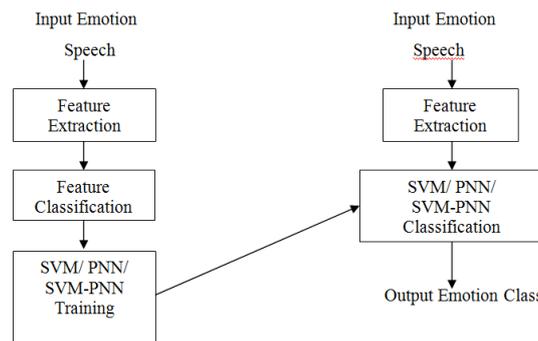


Fig. 3 Structure of Speech Emotion Recognition System

3.1 Input

The input is .wav format which is a emotional utterance from database. Two database is created i.e training database and training database. The database expresses five emotional states Surprise, Sad, Neutral, Happy, Fear is used for conducting the experiment. The base material of the database is consist of clips that is extracted from the selected recording. An audio .wav format file edited to remove sounds other than the main speaker. Some samples used for training purpose and some samples are used for testing for testing purpose.

3.2 Feature Extraction

In feature extraction process extracts the feature from speech samples which contains maximum information related to human emotions. A proper selection of features can increase the efficiency of the system. In this proposed system two features are used i.e MFCC and Energy for extracting purpose. The steps for calculating the MFCC feature are shown above.

3.3 Classifier

A. Support Vector Machine

The support vector machine is classifier used as a classification purpose for speech recognition. SVM is a algorithm used in pattern recognition for regression and data classification[3]. The classifier is used for classification purpose or separating the features from other feature. SVM perform classification by constructing N- dimensional hyper-plane which seperates the data into categories. A good separation is achevied by creating two support vectors with largest distance to the nearest training data point of any class called as functional margine. The classification is achieved by non-linear or linear separating surface in the input space. SVM is binary classifier but with some approaches it is used for multiclass classifier. Common two methods builds binary classifier are where each classifier distinguishes between (i) one of the labels to the rest (one-versus-all) or (ii) between every pair of classes (one-versus-one). Classification of new instance for one –verses- all case is done by winner takes- all strategy, in which classifier with highest output function assign the class. One-verses-one classification is done by using max-wins voting strategy, in which every classifier allot the instance to one of the two class, then vote for the allotted class is increased by one vote. At last the class with most votes determine the instances classification[9].The Fig. 4 for one-versus-all and Fig. 5 for one-versus-one is shown below.

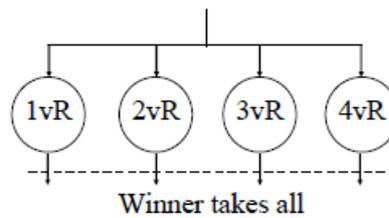


Fig. 4 One-versus-all Approach

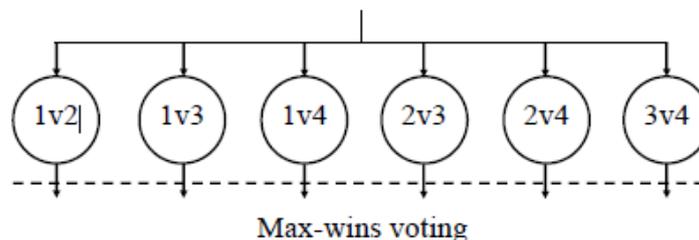


Fig. 5 One-versus-one Approach

In this paper classification is carry out with assist of Multiclass classifier i.e. one-to-one multiclass approach, in which each classifier assign the instance to one of the two classes, then the vote for the assign class is increased by one vote, and at last the class with the most votes determine the instance classification. The one-versus-one (1v1) classifier uses a “max-wins” voting strategy[20]. It construct $m(m - 1)/2$ binary classifiers, one for each pair of different classes. Every binary classifier C_{ij} is trained on the data from the i th and j th classes only. For a known test sample, if classifier C_{ij} predict it is in class i , then the vote for class i is increased by one; or else the vote for class j is increased by one. Then the “max-wins” voting strategy assign the test sample to the highest scoring class.

B. Probablistics Neural Network (PNN)

The Probablistics Neural Network was first proposed by Specht . With adequate training data, the PNN is sure to converge to a Bayesian classifier, and thus, it has a great probable for making classification decisions correctly and providing reliability and probability measures for every classification. In addition, the training procedure of the PNN only desires one epoch to adjust the biases and weights of the network architecture[8]. Therefore, the most important advantage of using the PNN is its high speed of learning. In general, the PNN consists of an input layer, a pattern layer, a summation layer, and a decision layer as shown in Fig. 6.

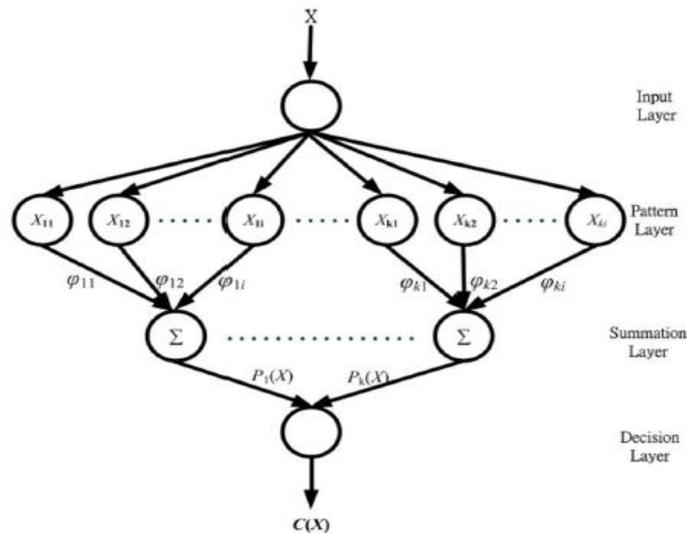


Fig. 6 Topology of a PNN classifier.

The purpose of the neurons in each layer of the PNN is defined as follows.

- Layer 1:** The first layer is the input layer, and this layer perform no computation. The neurons of this layer communicate the input features \mathbf{x} to the neurons of the second layer directly

$$\mathbf{x} = [x_1, x_2, \dots, x_p]^T$$

where p is the number of the extracted features.

- Layer 2:** The second layer is the pattern layer, and the number of neurons in this layer is equivalent to NL . Once a pattern vector \mathbf{x} from the input layer arrives, the output of the neurons of the pattern layer can be calculate as follows:

$$\varphi_{ki}(x) = \frac{1}{(2\pi)^{d/2}\sigma^d} \exp\left(-\frac{(x - x_{ki})^T(x - x_{ki})}{2\sigma^2}\right)$$

where x_{ki} is the neuron vector, σ is a smoothing parameter, d is the dimension of the pattern vector \mathbf{x} , and ϕ_{ki} is the output of the pattern layer.

- Layer 3:** The third layer is the summation layer. The contributions for each class of inputs are summed in this layer to generate the output as the vector of probabilities. Each neuron in the summation layer represent the dynamic status of one class. The output of the k th neuron is

$$p_k(x) = \frac{1}{2\pi^{d/2}\sigma^d} \frac{1}{N_i} \exp\left(-\frac{(x - x_{ki})^T(x - x_{ki})}{2\sigma^2}\right)$$

where N_i is the total number of samples in the k th neuron.

- Layer 4:** The fourth layer is the decision layer

$$c(\mathbf{x}) = \arg \max \{p_k(\mathbf{x})\}, \quad k = 1, 2, \dots, m$$

where m denotes the number of classes in the training samples and $c(\mathbf{x})$ is the estimated class of the pattern \mathbf{x} . If the *a priori* probabilities and the losses of misclassification for each class are all the same, the pattern \mathbf{x} can be classified according to the Bayes' plan in the decision layer based on the output of all neurons in the summation layer[20].

Hybrid Method (SVM-PNN)

In this paper, a combination of Support vector Machine (SVM) and Probabilistic neural network (PNN) is proposed. This paper clear that the result can be improve by combining the properties of SVM and PNN. The most important advantage of using the PNN is its high speed of learning. PNN systems can be applied to multi-class classification problems in a natural way. In general, the PNN consists of an input layer, a pattern layer, a summation layer, and a decision layer. Based on this four layer it can classify the speech samples. Support Vector Machine performs classification by constructing an N-dimensional hyper-plane that optimally separates the data into categories. It is one of the best methods for classification of emotions from speech. Hence we proposed a system based on combination of PNN and SVM. A comparative analysis of result of single PNN, single SVM and hybrid model of both is done.

IV. RESULTS

Table I Confusion matrix for support vector machine

Emotions	Surprise	Sad	Neutral	Happy	Fear
Surprise	2	0	0	0	0
Sad	1	1	0	0	0
Neutral	0	0	2	0	0

Happy	1	0	0	1	0
Fear	0	0	1	0	1

Accuracy of SVM=70%

Table II Confusion matrix for probabilistic neural network

Emotions	Surprrise	Sad	Neutral	Happy	Fear
Surprrise	1	0	1	0	0
Sad	1	1	0	0	0
Neutral	1	0	1	0	0
Happy	1	0	0	0	1
Fear	0	0	1	0	1

Accuracy of PNN=40%

Table III Confusion matrix for svm-pnn hybrid classifier

Emotions	Surprrise	Sad	Neutral	Happy	Fear
Surprrise	2	0	0	0	0
Sad	1	1	0	0	0
Neutral	0	0	2	0	0
Happy	0	0	0	2	0
Fear	0	0	0	0	2

Accuracy of SVM-PNN Hybrid Classifier=90%

For comparison of Results the three different approaches as PNN, SVM, SVM-PNN represented in graphical form is as shown in figure. The result obtain using hybrid combination is comparatively superior from remaining two methods. The overall accuracy of proposed method is 90%.

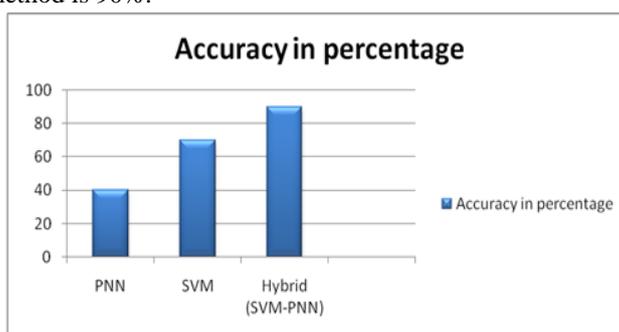


Fig. 7 Overall Accuracy Of Hybrid Model

IV. CONCLUSION

Emotional Recognition System is current research topic with wide range. Human Emotion Recognition System have many application in the field of computer and Human Interaction for improvement in communication between Human and Machine Interaction. In day to day Life, Speech Emotion Recognition System is gaining lot of significance.

In this proposed system feature extracted by using Energy and MFCC from Emotional Database. The Human Emotion Recognition System's Accuracy by using PNN is 40% and by using SVM is 70%. In proposed system Hybrid Model i.e SVM-PNN is designed and implemented whose accuracy is 90% which is better than single SVM and single PNN.

ACKNOWLEDGEMENT

I would like to say thanks to my guide "Prof. Surendra Waghmare" who gave his knowledge and time in order to complete this paper. This paper would never complete without his and the support of faculty members.

REFERENCES

- [1] Albino Nogueiras, Asunción Moreno, Antonio Bonafonte, and José B. Mariño, "Speech Emotion Recognition Using Hidden Markov Models", Eurospeech 2001 – Scandinavia.
- [2] Dimitrios Ververidis, Constantine Kotropoulos, "Emotional speech recognition: Resources, features, and methods", 2006 Elsevier
- [3] Sujata B. Wankhade, Pritish Tijare, Yashpalsing Chavhan, "Speech Emotion Recognition System Using SVM AND LIBSVM", International Journal Of Computer Science And Applications Vol. 4, No. 2, June July 2011, ISSN: 0974-1003
- [4] Prof. Sujata Pathak, Prof. Arun Kulkarni, "Recognizing emotions from Speech", 978-1-4244-8679-3/11/\$26.00, 2011 IEEE.....block diagram
- [5] Marcin Blachnik, Włodzisław Duch, "LVQ algorithm with instance weighting for generation of prototype-based rules", 2011 Elsevier Ltd.

- [6] M.N.Hasrul, M.Hariharan, Sazali Yaacob,” Human Affective (Emotion) Behaviour Analysis using Speech Signals: A Review”, 2012 International Conference on Biomedical Engineering (ICoBE),27-28 February 2012, Penang
- [7] Lijiang Chen , Xia Mao, Yuli Xue , Lee Lung Cheng ,” Speech emotion recognition: Features and classification models”, 2012 Elsevier
- [8] Jeen-Shing Wang, *Member, IEEE*, and Fang-Chen Chuang,”An Accelerometer-Based Digital Pen With a Trajectory Recognition Algorithm for Handwritten Digit and Gesture Recognition”, *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*, VOL. 59, NO. 7, JULY 2012.
- [9] Shashidhar G. Koolagudi · K. Sreenivasa Rao,” Emotion recognition from speech: a review”, *Int J Speech Technol* (2012) 15:99–117 DOI 10.1007/s10772-011-9125-1.
- [10] Vaishali M. Chavan, V.V. Gohokar,“Speech Emotion Recognition by using SVM-Classifer”, *International Journal of Engineering and Advanced Technology (IJEAT)* ISSN: 2249 – 8958, Volume-1, Issue-5, June 2012.
- [11] Yixiong Pan, Peipei Shen and Liping Shen,”Speech Emotion Recognition Using Support Vector Machine”, *International Journal of Smart Home* Vol. 6, No. 2, April, 2012.
- [12] Vaishali M. Chavan, V.V. Gohokar,” Speech Emotion Recognition by using SVM-Classifer”, *International Journal of Engineering and Advanced Technology (IJEAT)* ISSN: 2249 – 8958, Volume-1, Issue-5, June 2012
- [13] Bhoomika Panda , Debananda Padhi, Kshamamayee Dash, Prof. Sanghamitra Mohanty,”Use of SVM Classifier & MFCC in Speech Emotion Recognition System”, Volume 2, Issue 3, March 2012 ISSN: 2277 128X *International Journal of Advanced Research in Computer Science and Software Engineering*
- [14] Dipti D. Joshi, Prof. M.B. Zalte,“ Speech Emotion Recognition:A Review ”,*IOSR International Journal Of Electronics and Communication Engineering(IOSR -JECE)*, ISSN 2278-2834,ISBN:2278-8735,Volume 4, Issue4,PP 34-37.(Jan-Feb. 2013).
- [15] Thapanee Seehapoch, Sartra Wongthanavasuu,” Speech Emotion Recognition Using Support Vector Machines”, 2013 5th International Conference on Knowledge and Smart Technology (KST)
- [16] Mohammad Masoud Javidi and Ebrahim Fazlizadeh Roshan,” Speech Emotion Recognition by Using Combinations of C5.0, Neural Network (NN), and Support Vector Machines (SVM) Classification Methods”, *Journal of mathematics and computer Science* 6 (2013), 191-200
- [17] Milton, S. Sharmy Roy, S. Tamil Selvi, PhD” SVM Scheme for Speech Emotion Recognition using MFCC Feature” , *International Journal of Computer Applications (0975 – 8887) Volume 69– No.9, May 2013*
- [18] A. Milton, S. Tamil Selvi,” Class-specific multiple classifiers scheme to recognize emotions from speech signals”, 2013 Published by Elsevier Ltd
- [19] Nermine Ahmed Hendy and Hania Farag,” Emotion Recognition Using Neural Network: A Comparative Study”, *World Academy of Science, Engineering and Technology* Vol:7 2013-03-20
- [20] Shubhangi S.Jarande, Prof. Surendra Waghmare,” A Survey on Different Classifier in Speech Recognition Techniques.”, *International Journal of Emerging Technology and Advanced Engineering*, (ISSN 2250- 2459,ISO 9001:2008 Certified Journal, Volume 5, Issue 3, March 2014)