Back Propagation Neural Network based Emotion Recognition System

Rohit Katyal[1]

[1,2] University Institute of Engineering & Technology Chandigarh University
Mohali, India

***Abstract*- Emotion recognition with speech has now a day's getting attention of engineers in field of pattern recognition and speech signal processing. As computers is one of the important part now a days, so the requirement for communication between computer and humans. Through voice signals Automatic emotion recognition recognise the emotional state of the speaker. In humans life emotions plays a vital role. This paper focuses on the e motion recognition process by including feature extraction of the speech signals and then classification process with BPNN Classifier. In this system, a database of 500 speech signals are used to train and test the system which are of different categories like happy, sad, aggressive and fear. Classification process is depending on different extracted features like Maximum Frequency, Average Frequency, Minimum Frequency, Roll off, Pitch, and Loudness from a speech sample. The classification accuracy achieved by this system varies from 85 to 95 percent.**

Keywords— **BPNN, Emotion Recognition, Speech Recognition**

I. INTRODUCTION

The most natural & fast communication method between the human is through speech signals. So speech is the fast & efficient method for interaction between humans & machines as well. To recognize the human voice machine should be intelligent machine so that it can also identify the emotions of the human voice known by speech emotion recognition. It can be used to extract useful semantics from the speech & improve their performance for speech recognition system.

Speech Emotion recognition is quite useful for numerous applications that requires man-machine interaction like in-car board system where for driver safety his/her mental state can be extracted by using voice signals. For therapists it works as a diagnostic tool. It may also be useful in the communication between parties where emotional state of speaker plays a major role also called automatic translation system. The other important applications are in some of call centre applications & in mobile communication.

Speech Emotion recognition is a very challenging task because of the following reasons. First, the concept of selecting feature for emotion recognition is not clear because there is no prior information about the features that recognize emotions. The most common features like pitch & energy contours can be affected by existence of different speakers, speaking styles, speaking rates & sentences. The second issue is different expressing way because it generally depends on the speaker, their culture & environment**.** Other problem which may be that one might undergo some emotional state like sadness for weeks, days or months. In such a case, emotions can be transient and cannot be detected.

Emotional speech recognition points at identifying physical and emotional condition involuntarily of humans using her or his voice. Speaker does not have similar stages throughout the speech which are to recognise as speeches emotional aspect and that are integrated in paralinguistic aspect. The linguistic content cannot modify by emotional state; in communication of individual this is a significant factor, since feedback information is provided in numerous applications. Speech is possibly the generally proficient way to correspond with each other [18]. This too means that speech could be a helpful boundary to cooperate with machines. A few victorious examples based on it throughout the past years, while we have awareness about electromagnetism; includes the development of the megaphone, telephone. Even in the previous centuries people were researching on speech fusion. Von Kempelen developed an engine talented of 'speaking' words and phrases. At the present time, it has happened to achievable not only to expand examination and execute speech recognition systems, but also to have systems competent to real-time alteration of text into speech. Regrettably, in spite of the high-quality development made on that area, there are countless applications that are the speech recognition procedure facing dig now; speech is a very prejudiced experience that is added by the majority of them. It is very complex task to recognise the words from speech.

II. BACK PROPAGATION NEURAL NETWORK

Any Algorithm of supervised learning category has a goal of finding a function that maps a group of inputs with its output. For Example: a classification task, where animal is taken as image of an input and correct output is animal's name. Some output and input patterns can be learned easily by neural networks of single layer (i.e. preceptors). However some patterns cannot be learned by perceptrons of single layer like linear separable ones. For example, animal's image may be classified by a human by some features like skin texture, limbs number, animal size etc. However neural network of single

layer must learn the function that uses the intensity of image pixels to output label. As it is restricted to have a single layer, inputs abstract features cannot be learnt. A network of multi layer category can overcome the limitation as it can learn features of different kind and create internal representation in each and every layer. First layer is concerned with learning orientation of lines with the help of inputs provided by each and every pixel in image. Second layer concerned with identifying shapes like that of a circle, and combing learned features in first layer. Higher layers than these learn more abstract features like ones mentioned above which could be in image classification. A layer below every layer find patterns and it is because of this ability of creating internal representations which are free from outside input that gives power to networks of multi layered category. Algorithm of back propagating type was developed with the motivation and goal to finding a way of training neural network of multilayered type so that it could learn internal representations for allowing it to learning mapping of arbitrary type from input to output [1].

Back propagation is used in training artificial neural network which is used with optimization method like gradient descent. Loss function gradient is calculated by this method in regard all weights in network. Gradient uses optimization method for weight updates to minimize function loss.

Black propagation algorithm is divided into 2 phases: weight update and propagation.

Phase 1: propagation

Following steps are involved in each propagation:

1. To generate propagations output activations, forward propagation of training patterns input through neural network.

2. To generate deltas of hidden and output neurons, backward propagations of propagations output activations through neural network using training pattern.

Phase 2: weight update

Following steps are to be followed for every weight synapse:

1. Multiply input activation and output delta for getting weight gradient.
2. Subtract from weight ratio of gradient

Quality and speed of learning is influenced by this ratio; it is called learning rate. Training is accurate, if ratio is lower; if ratio is greater neurons trains faster. Gradient of weight sign tells about where error is increasing; that is why opposite direction should be used to update weight. Till the networks performance is satisfactory repeat 1 and 2 phase.

Rumelhart et al developed (BPNN) back propagation neural network [12] as an answer to a problem regarding Training of multi layer perceptrons. BPNN represented some fundamental advances:-

1. Utilization of error back propagation to change the weights of internal network after every training epoch.
2. At networks each node inclusion of differentiable transfer function.

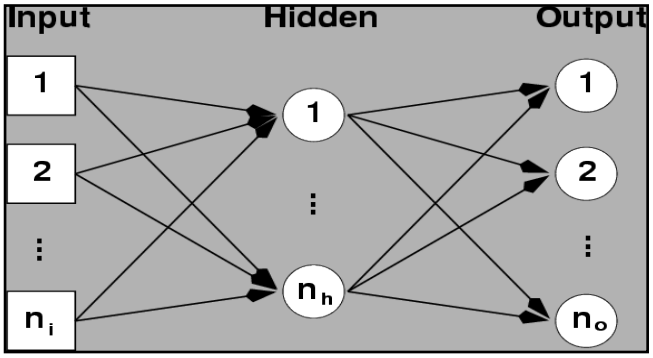

Fig. 1 Back Propagation Neural Network

Back propagation neural network is shown in schematic form in Fig 1. BPNN used in the work have three layers of nodes or neurons (output, hidden and input). Every node in hidden and input layers is connected with every node in next layer (output or hidden). Connections between each node is directed (i.e. one way flow of the information), and for nodes that are within particular layer there is no connection. For every connection that is present between nodes, has a typical weighting factor linked with it. Back propagation method is used to modify weights during process of training to yield 'learning'

Primarily BPNN was selected classifier because of its capacity of generating in future space complex decision boundaries [13]. There is also some work that tells BPNN, under correct situations can approx at its output probabilities of Bayesian posterior [14]. This is important as possibly the best performance (i.e. error rate which is lowest) is provided by Bayesian classifier for a provided distribution of data. For patterning classification by using different non parametric methods, it is difficult to guess performance of Back propagation neural network a priori. In addition there are various parameters of Back propagation neural network which must be selected, including learning rate, hidden notes numbers, and training sample numbers.

According to the work of Haussler and Baum [15], we can place bound (m) on training samples required to guarantee a required degree of performance on particular set of test samples produced from same old distribution as training data. Particularly ,if we use m samples for training with W weight networks and N nodes so that we get a fraction equal to(1-e/2) of them are produced correctly , then that means that one could be sure that a fraction (1-e) of future samples (test) from same old distribution would be classified rightly, where

$$m \geq O\left(\frac{W}{\in} \log \frac{N}{\in}\right)$$

As a particular example, to promise not more than 10% error while classifying test data .Training sample numbers should be roughly 10 times to number of weights in network. For a particular network which is generated below, it shows a need for training samples 5000-10000. It is not tractable to produce that number of images.

The basis for choosing number of hidden nodes to be utilised in single hidden layer network is not that developed. The only method to optimize this is by testing network with different numbers of hidden node and selecting the one whose performance is the best

III. PROPOSED WORK

The main and crucial step of Speech Recognition Systems is to recognise the speech effectively. Here word "effectively" means to recognise the words accurately on the basis of features extracted whether it is SAD, HAPPY etc. Recognizing the emotions in speech is a complex process. Complexity of this process is about that the environment or source that generates speech is difficult to identify, source may be male or female. This work is to identify the emotion of the speech based on four categories namely SAD, HAPPY, ANGRY and FEAR using BPA algorithm. The results of this system can be measured by implementing the whole system environment on MATLAB environment.

A. Basic Design

The accuracy of the emotion recognition system depends on the database to be used. In this work, a database of 500 speech samples can be used. The basic design of the system is shown in figure 1 which defines the whole process of this system.

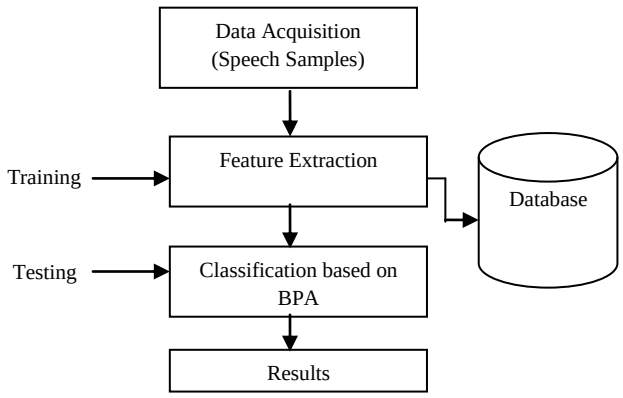

Fig 2: Basic Design of the System

This system involves the following steps:

a. Data Acquisition: This step involves the data acquisition process where a number of speech samples are collected. In this work 125 samples of each category can be acquired. The categories involve happy, sad, and aggressive and fear sound samples saved as a .wav files.

b. Feature Extraction: This step performs feature extraction process and save all the features in the database. The features extracted are Maximum Frequency, Average Frequency, Minimum Frequency, Roll off, Pitch, and Loudness etc. It is also known to as a training phase.

c. Feature Classification: This step also called as a testing phase. Here in this step all the features can be classified on the basis of BPNN classifier. Back propagation neural network (BPNN) is the for the most accepted and commonly used neural network [2-7] .Usually back-propagation neural network consists of feed forward multi-layer network in an input layer, an output layer, and at least one hidden layer. All layers are completely associated to the successive layer. There is no interconnection or feedback between PEs in the same level. Each PE has a predisposition input with non-zero weight. The bias input is analogous to the. In a neural network consisting of P processing elements, the input/output function is defined as:

$$k = L(mW)$$

Where m = ***{m,)*** is the input vector to the network, k = {Kkj} is the output vector from the network, W is the weight matrix. It is afterward defined as

$$W = (w_i^T, w_j^T, \ldots\ldots\ldots\ldots\ldots, W_n^T)^T$$

Where the vectors w_i ,wj ,...,Wn are the individual PE weight vectors, which are given as:

$$W_T = \begin{cases} w_i \\ w_j \\ \vdots \\ w_n \end{cases}$$

The main function of back propagation neural network is that it moves from input to output layer. In the back propagation neural network error is found at output layer, so to remove error at output layer, moves backward takes place, hence name is back propagation neural network.

B. Two Phase Design

This system works on two phases which involves training phase where feature extraction process can be done to prepare database and the other is the testing phase where different classifier can be used for detection of emotions from speech signal. These are explained as follows.

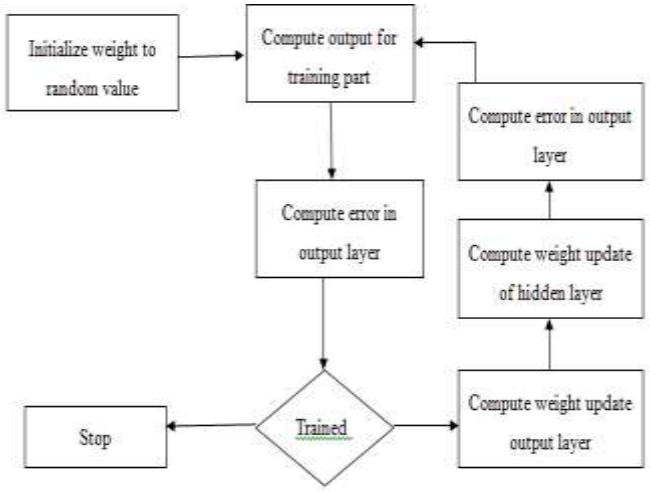

Fig 3: Back Propagation Training Algorithm

Training: The training section ensures that the database gets trained properly so that at the time of testing it produces wide-ranging results. The characteristics of the training are as follows.

- **Maximum Frequency:** It is the rate which we get at the tip on a frequency chart. When so ever we place a voice sample over the time and frequency model, the maximum peak is known as the maximum frequency of the voice illustration.
- **Minimum Frequency**: It is the rate which we get at the tip on a frequency chart. When so ever we place a voice sample over the time and frequency model, the minimum peak is known as minimum frequency of the voice illustration.
- **Average Frequency**: The average frequency can be considered using two techniques. The first technique is to append all the frequency samples and then separate the entire sum with the total number of frequency. The second method is an incredibly principled method in which we can insert the minimum frequency and the maximum frequency and then we can divide them by two.

$$Average\ Frequency = (Minimum\ frequency + Maximum\ Frequency)/2$$

- **Spectral Roll off**: The spectral roll off is difference between the maximum frequency differences with the adjacent frequency.
- **Noise Level:** The noise level is the additional amount of bits which has been supplementary into the voice sample.
    1. **Uniform Noise**: Uniform noise is the noise which is concurrently equivalent all over the voice sample.
    2. **Non Uniform Noise**: Non-uniform noise is the noise that does not remain constant all over the sample.
- **Pitch**: It is the standard value of the entire voice sample.
- **Spectral Frequency**: The spectral frequency is the frequency of the voice pitch subsequently to the highest voice sample.

Testing: In the testing phase, the trained network was computer-generated with unknown speech blueprint. It was observed that the trained network performs very well and more than ten words can be recognized by using the developed system [14].

IV. Results & Discussions

The results of the proposed system can be analyzed by implementing the whole methodology using simulation tool MATLAB. The experiment has been done on the basis of four speech categories namely sad, happy, angry and aggressive in MATLAB 2010. The processing has been captured & then accuracy can be calculated by using BPNN. BPNN takes the extracted features as an input & then testing their accuracy on the basis of different parameters defined below.

a) Epochs: Epochs is the number of iterations which the neural network runs. In this, there are three iterations out of 500 which have successfully predicted the category of the speech file. If the system runs 3 iterations, it does not predict that the best result would come out at third iteration though the network takes an average of the three iterations and depicts the best results out of time.

b) Time: It determines the total time consumed in the prediction.

c) Performance Measure: It represents the elapsed performance hit in the network architecture which can be plotted through the SOM plots of the neural network.

d) Validation Checks: It depicts the number of validations which the network can apply to the testing procedure.Table 1 gives the percentage accuracy achieved while testing the signals of different categories.

Table 1: Accuracy percentage of SAD, HAPPY, ANGRY, AGGRESIVE moods

CATEGORY	SAD	HAPPY	ANGRY	AGGRESSIVE
SAD	93	3	2	2
HAPPY	2	94	3	1
ANGRY	1	1	95	3
AGGRESSIVE	2	2	3	93

The graph represented below shows accuracy percentage of mood swings like SAD, HAPPY, ANGRY, AGGRESIVE.

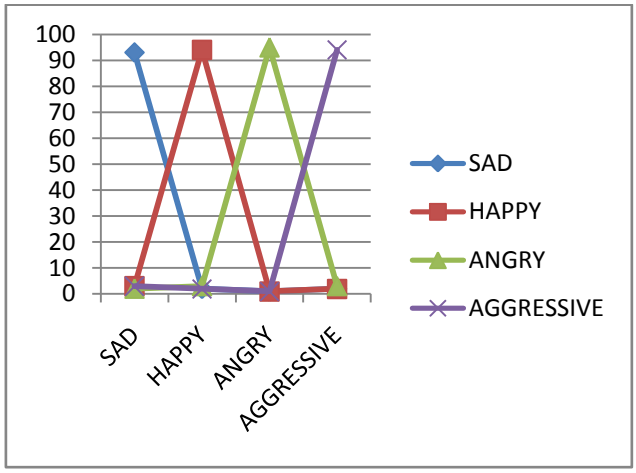

Fig 4: Graph representing BPNN accuracy

V. CONCLUSION & FUTURE SCOPE

The most natural & fast communication method between the human is through speech signals. So speech is the fast & efficient method for interaction between humans & machines as well. In this work, for speech emotion recognition-BPA classifier is utilised in selected speech data . Every classifier has a separate theory for its implementation. From all the calculations done above we come to a conclusion that using BPA(Back propagation algorithm) for speech emotion detection belonging to neural network performs better and with high accuracy. The classification accuracy for BPA algorithm varies from 85 to 95 percent. The current research work opens a lot of gates for the future research workers. This research work is limited for the users that mean the current system has to be trained for the every user you want to identify. The other aspect of this research work is that it does not provide any gender information in terms of the classified data set that means it does not specify that the classified voice is for male or for female.

REFERENCES

[1] B. Jaggi, S. S. Poon, C. MacAulay, and B. Palcic, ``Imaging system for morphometric assessment of absorption or fluorescence in stained cells," *Cytometry*, vol. 9, no. 6, pp. 566-572, 1988.

[2[] M. A. Hossain, M. M. Rahman, U. K. Prodhan and M. F. Khan, "Implementation Of Back-Propagation Neural Network For Isolated Bangla Speech Recognition," *International Journal of Information Sciences and Techniques* , Vol 3, No.4, pp. 1-9, 2013.

[3] R. Rojas," Neural Networks:A Systematic Introduction," *Springer Berlin Heidelberg New York*, 2005.

[4] Ververidis and C. Kotropoulos "Emotional speech recognition: Resources, features,and methods," *Speech Communication* , Vol 48, pp. 1162-1181

[5] J. C. Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines", *Advances in Neural Information Processing Systems* 11, M. S. Kearns, S. A. Solla, D. A. Cohn, eds., MIT Press, (1999).

[6] W. Gevaert, G. Tsenov and V. Mladenov, "Neural Networks used for Speech Recognition," *Journal of Automatic Control*, Vol. 20, pp. 1-7, 2010.

[7] M. Cilimkovic, "Neural Networks and Back Propagation Algorithm", Institute of Technology Blanchardstown, Blanchardstown Road North Dublin 15, Ireland.

[8] P. Peng, Q. L. Ma and L. Hong, "The Research Of The Parallel Smo Algorithm For Solving Svm," *Proceedings of the Eighth International Conference on Machine Learning and Cybernetics, Baoding*, 2009.

[9] R. Fan, P. Chen and C. Lin, "Working Set Selection Using Second Order Information for Training Support Vector Machines," *Journal of Machine Learning Research* Vol. 6 , pp. 1889–1918, 2005.

[10] X. Shao, KunWu, and B. Liao, "Single Directional SMO Algorithm for Least Squares Support Vector Machines," *Computational Intelligence and Neuroscience* Vol., 2013, Article ID 968438.

[11] F. R. Bach, G. R. G. Lanckriet and M. I. Jordan, "Multiple Kernel Learning, Conic Duality, and the SMO Algorithm", *Proceedings of the 21st International Conference on Machine Learning*, 2004.

[12] D. Rumelhart, G. Hinton, and Williams R., ``Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533-536, 1986.

[13] Kurt Hornick, Maxwell Stinchcombe, and Halbert White, ``Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359-366, 1989.

[14] Michael D. Richard and Richard P. Lippmann, ``Neural network classifiers estimate bayesian a posteriori probabilities," *Neural Computation*, vol. 3, no. 4, pp. 461-483, 1991.

[15] E.B. Baum and D. Haussler, ``What size net gives valid generalization?," *Neural Computation*, vol. 1, no. 1, pp. 151-160, 1989.

[16] N.Pushpa, R.Revathi, C.Ramya and S.Shahul Hameed, "Speech Processing Of Tamil Language with Back Propagation Neural Network and Semi-Supervised Training," *International Journal of Innovative Research in Computer and Communication Engineering*, Vol.2 (1), 2014.

[17] H. Gupta and D. S. Wadhwa, "Speech Feature Extraction and Recognition Using Genetic Algorithm," *International Journal of Emerging Technology and Advanced Engineering*, Vol. 4 (1), 2014.

[18] J. Bachorowski, "Vocal Expression and Perception of Emotion," *Current Directions in Psychological Science*, Vol. 8 (2), pp. 53–57, 1999.

[19] D. A. Sauter; F. Eisner, A. J. Calder and S. K. Scott, "Perceptual cues in nonverbal vocal expressions of emotion," *The Quarterly Journal of Experimental Psychology*, Vol. 63 (11), pp. 2251–2272.

[20] Y. Pan, P. Shen , and Liping Shen," Speech Emotion Recognition Using Support Vector Machine" , International Journal of Smart Home Vol. 6, No. 2, April, 2012.

[21] A. Utane, S.L Nalbalwar, "Emotion Recognition Through Speech Using Gaussian Mixture Model And Hidden Markov Model", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013.