FPGA Implementation of Emotion Recognition from Speech Signals

N. Sowjanya¹, B. Rajasekhar², M. Kamaraju³
M. Tech student ¹, Associate Professor², Professor & HOD³
Gudlavalleru Engineering College
Gudlavalleru, Andhra Pradesh, India

Abstract— In this paper speech based emotion recognition based on ANNs is presented. The main objective of the project work is, given the speech signal as input to recognize the Gender (male or female) and emotion (happy, sad, angry, and lazy) of the speaker the extraction of emotion from speech signal is done by considering emotional speech database. Taken the speech signal from emotional speech database the features like pitch, energy, frequency coefficients of the speaker were extracted. After the feature extraction block we have post processing in which the speech signal containing noise and the dimension of the speech signal is decreases. The most important step in speech emotion recognition system is choosing the best classifier for classification of emotion. The performance of the system is based on choosing the best classifier. The existing classifiers are HMM, GMVAR, SVM, and LDA. In this paper back propagation algorithm of the NNs is used. The results obtained were improved in device utilization and speed than the existing method like LDA.

Keywords----Artificial Neural Networks (ANNs), Latent Dirichlet Allocation (LDA), Hidden Markov Model (HMM), Gaussian Mixture Vector Auto Regressive (GMVAR) and Support Vector Machine (SVM).

I INTRODUCTION
There are many ways of communicating between the two human beings but speech has proven to be a good indicator of emotional content. In speech based emotion recognition we have to give the input speech signal from that we have to find out the emotions of the speaker. The emotions are happy, sad, angry, and lazy of the speaker. Speech [1] is the best indicator to recognize the emotions of the speaker. In speech based emotion recognition system the main thing is to choose the algorithms to represent the emotions with high accuracy and allow implementation complexity. The speech based emotion recognition system has many applications in areas like medical, security and surveillance, personnel memory aids life logs [2 3].

Typical speech based emotion recognition is shown in Fig. 1. In block diagram there are mainly four blocks namely

1. Feature extraction
2. Post processing
3. Modeling
4. Recognition

In feature extraction we have to extract the features like energy, pitch, frequency coefficients of the speaker. To provide the good representation of emotional content [4] in speech signal we have to consider Activation, Valence, and Dominance.

Fig. 1 Speech emotion recognition system

Generally post processing is used for dimensionality reduction of the speech signal. The speech based emotion recognition is based on choosing the best classifier. Choosing of best classifier it gives the best results in emotion recognition frame work. The different classifiers are existed such as Hidden Markov Model (HMM) [5] and Gaussian Mixture Vector Auto Regressive models (GMVAR) [6]. Latent Dirichlet Allocation (LDA) model is a generative probabilistic model for document preparation. In this we can create one document. Every document is divided into topics and every topic is divided words and every word is divided into vocabulary. In speech emotion recognition frame work speech signal is taken as one document and in that topics are Gender recognition and Emotion recognition. In that words

137 | N. Sowjanya, B. Rajasekhar, M. Kamaraju
are male and female in Gender recognition and in that happy, sad, angry, and lazy in Emotion recognition system.

II ARTIFICIAL NEURAL NETWORKS (ANN)

Artificial Neural Networks are same as that of biological nervous system of animals and also motivated by the human computing system in particularly the brain. Different learning mechanisms are available in Artificial Neural Networks to enable the neural network to acquire the knowledge. Based on the learning mechanisms neural networks are classified into different types. The learning process in neural networks is called training [19] and the ability to solve the problem is called inference.

ANNs are simplified imitations of the central nervous system, and obviously therefore, have been motivated by the computing performed by human brain. The human brain is a highly complex structure viewed as a massive, interconnected network of processing elements called neurons. However, the behavior of a neuron can be captured by a simple model as shown in Fig. 2.

![Simple model of ANN neuron](image)

Fig. 2 Simple model of ANN neuron

Here, X₁, X₂, X₃ are the n inputs to the artificial neuron. Here w₁, w₂, w₃……w n are the weights attached to the input links.

A biological neuron sum the inputs received through the dendrites, and produce an output if the sum is greater than threshold value. The input signals are passed on to the cell body. If the sum is less than a threshold value it should not pass the signal through the cell body. An effective synapse which transmits the stronger signal will have a large weight while a weak synapse will have smaller weights. Here weights will have multiplication factor of the inputs to account for the strength of synapse. The total input I received by the sum of the artificial neuron is

\[ I = w₁X₁ + w₂X₂ + w₃X₃ + \ldots + wₙXₙ \]

III PROPOSED WORK

The most important step in emotion recognition system through speech is choosing the best classifier. The performance of the speech emotion recognition system is depending on the classifier. The classification of emotion can be done by giving different features to the classifier.

In the proposed method Artificial Neural Networks (ANNs) are used as a classifier. To train the network according to our requirement we must one algorithm for this purpose we have used the Artificial Neural Networks (ANNs) of the Back Propagation algorithm t is used.

The steps for the back propagation algorithm follow as.

Step 1: The input layer inputs \{I\}_₁ and output layer outputs \{O\}_₀ are represented using “l” and “n” respectively.

Step 2: The number of hidden layer neurons between

\[ 1 < m < 2l \]

Step 3: The weights connecting to input and hidden neurons are represented as \[v\] and \[w\] represents the weights connecting to hidden layer and output layer neurons.

Step 4: By using the linear activation function the input layer input is same as that of input layer output. \{O\}_₁=\{I\}_₁

Step 5: Compute hidden layer input by multiplying the weights to the input layer output.

\[ \{I\}_H=[V]^T\{O\}_1 \]

Step 6: The hidden layer outputs are calculated using the sigmoidal function.

\[ \{O\}_H = \frac{1}{1+e^{-(\{I\}_H)}} \quad (1) \]

Step7: Compute output layer input by multiplying corresponding weights to the hidden layer output.

\[ \{I\}_O=[W]^T\{O\}_H \]

Step8: The output layer output is find out by using sigmoid function as

\[ \{O\}_O = \frac{1}{1/(1+e^{-(\{I\}_O)})} \quad (2) \]

Here the equation 2 is the network output.
Step 9: The difference between the Back propagation network output and the target output is the error for the \(i\)th training set as follows:

\[
E^p = \frac{1}{n} \sum (T_{oj} - o_{oj})^2
\]

Step 10: Find the \(\{d\}\) value as

\[
\{d\} = \{ (T_k - O_{ok}) O_{ok} (1-O_{ok}) \}
\]

Step 11: The \(Y\) matrix is calculated by multiplying the hidden layer output with \(d\).

\[
[Y] = \{O\}_H \{d\}
\]

Step 12: Find the updated weights as

\[
[\Delta W]^{t+1} = \alpha [\Delta W]^t + \eta [Y]
\]

Step 13: Calculate the error as \(\{e\} = [w] \{d\}\)

\[
\{d^*\} = \{ e_1 (O_{Hi}) (1-O_{Hi}) \}
\]

Step 14: Compute the \([X]\) matrix as by multiplying the \(d\) with input layer output.

\[
[X] = \{O\}_I \{d^*\} = \{I\}_I \{d^*\}
\]

Step 15: Find updated weights for input to hidden neurons and hidden to output neurons as follows

\[
[\Delta W]^{t+1} = [W]^t + [\Delta W]^{t+1}
\]

\[
[\Delta V]^{t+1} = [V]^t + [\Delta V]^{t+1}
\]

Step 16: Calculate the error rate as

\[
\text{Error rate} = \frac{\sum E_p}{n_{set}}
\]

Step 17: Repeat steps 4-16 up to the error will get minimized. 

End Back propagation algorithm.

**IV RESULTS**

The RTL schematic of speech emotion recognition system is as shown in Fig. 3. In this the inputs are 8 bit data, clock, reset, load and the outputs are female, male, angry, happy, sad, and lazy of the speaker.

![Fig. 3 RTL schematic of speech emotion recognition system](image)

The expanded view of RTL schematic is as shown in Fig. 4. RTL schematic is the top level view of our design. RTL schematic consists of basic gates (AND, OR), adders, multipliers.

![Fig. 4 RTL schematic of speech emotion recognition system](image)

The technology schematic of speech emotion recognition system is as shown in Fig. 5. Technology schematic consists of logic elements optimized for Xilinx device.

![Fig. 5 Technology schematic of speech emotion recognition system](image)

Simulation results of speech emotion recognition system are as shown in Fig. 6. The output signal is the male voice with angry emotion.

![Fig. 6 Simulation results of speech emotion recognition system](image)

Simulation results of speech emotion recognition system are as shown in Fig. 6. The output signal is the female voice with lazy emotion.
Simulation results of speech emotion recognition system are as shown in Fig. 6. The output signal is the female voice with angry emotion.

V CONCLUSION

In speech emotion recognition system the extraction of speech signal is taken from emotional speech database. By using the database, Gender (male or female) and emotions (Happy, Angry, Sad, and lazy) of the speaker were recognized. The area utilisation with the back propagation algorithm is less compared to LDA. The delay decreases with the ANN compared to LDA.

REFERENCES


