Speech Recognition using Neuro Fuzzy Network

Maitri Shah [1], Yash Sharma [2]

[1] *P.G student, Communication system engineering, SAL institute of technology and engineering research, Ahmedabad, Gujarat, India*
[2] *Assistant professor, E.C department, SAL institute of technology and engineering research, Ahmedabad, Gujarat, India*

ABSTRACT

Many popular algorithms are used to recognize the human voice. The good algorithms results high accuracy in recognition and robustness to the noises. In this research the performance of Neuro fuzzy system is verified for human voice recognition. Different set of words in Gujarati and English language are recorded in this research and are used as data sets in this experiment which can be applied to some instrument for different applications in future. Thus, this experiment shows that the Neuro fuzzy system is quiet robust to noise and can yield higher recognition results compared to other popular algorithms. Although speech recognition using Neuro fuzzy products are already available in the market at present their development is mainly based on statistical techniques which work under very specific assumptions. We elaborate the feasibility of an alternative approach for solving the problems more efficiently.

Keyword: - *Speech Recognition, Neuro Fuzzy Networks , Neural Networks, Fuzzy Logic, etc.*

1. INTRODUCTION

Speech recognition is a system in which the human speech or voice clip or any audio signal is recognized by the computers through the computational linguistics using different algorithms in the form of text. It is inter disciplinary sub field of computational linguistics that incorporates the knowledge and research in linguistics fields to develop technologies that enables the recognition and translation of spoken language into text. It is computer processed text to audio signal. Different techniques are applied for the speech recognition for analysis and synthesis of language and speech. Many algorithms are supportive to speech recognition like neural networks, fuzzy logics, discrete hidden markov model, fuzzy neural network, Neuro fuzzy networks etc. Artificial neural network is an information processing paradigm that is inspired by the way biological nervous system, such as brain process information. It has two modes of operation; The Training mode, where the element is trained to fire or not for particular input patterns and the Using mode where, a taught input method is detected at the input and its associate output becomes the current output. Fuzzy logic is used in computer technology to artificial intelligence.it demands linguistic rules in which the output and input variables has to be described linguistically and if the knowledge is wrong or contradictory fuzzy must be tuned. It has simple interpretations and implementation to the data sets. Neuro fuzzy system is trained by the means of a data driven learning method derived from neural network theory. It takes into account local information to cause local changes in fundamental fuzzy system

1.1 Method of Speech Recognition using Neuro-Fuzzy.

Neuro fuzzy system is trained by the means of a data driven learning method derived from neural network theory. It takes into account local information to cause local changes in fundamental fuzzy system. It is the system that is initialized with or without prior knowledge about the data. It approximates an n-dimensional unknown function which is partly represented by training examples. Fuzzy rules are interpreted as vague prototypes of the training data. Neuro fuzzy is represented as three layered feed forward neural network where first layer comprises of input variables, the second symbolizes fuzzy rules and the third layer gives us the output layer. Here the fuzzy sets are

converted to connection weights. This work use the Neuro-Fuzzy algorithm from to adapt for human sound because this algorithm has high accuracy for classifying infant sound when compare with other popular algorithm in classification. This algorithm has 3 sub-processes in training to creating the best recognition model and 1 sub process in testing. The training processes, including neural network, data normalization and fuzzy logic. The K-nearest Neighbors algorithm is used in the testing process. The first process is the neural network. The multilayer perceptron (MLP) with back-propagation technique is used. This process uses the data set from feature extraction part with 20 feature vectors. The Neural network structure is compound 3 parts.The structure of the neural network for each word including input layer, two hidden layers and output layer. The number of nodes in each layer are 20, 3, 2 and 1 respectively. The input layer receives the feature vectors and forward to the next layer. All weights and bias in hidden layers and output layer are randomly initialized. The structure of each node in the hidden layer and output layer which used in this work is shown. The output from each node calculate the summation of data (x), weight (w)and bias(α) from

$$y_i = \sum_{i=1}^{n} w_i x_i + \alpha$$

T

T

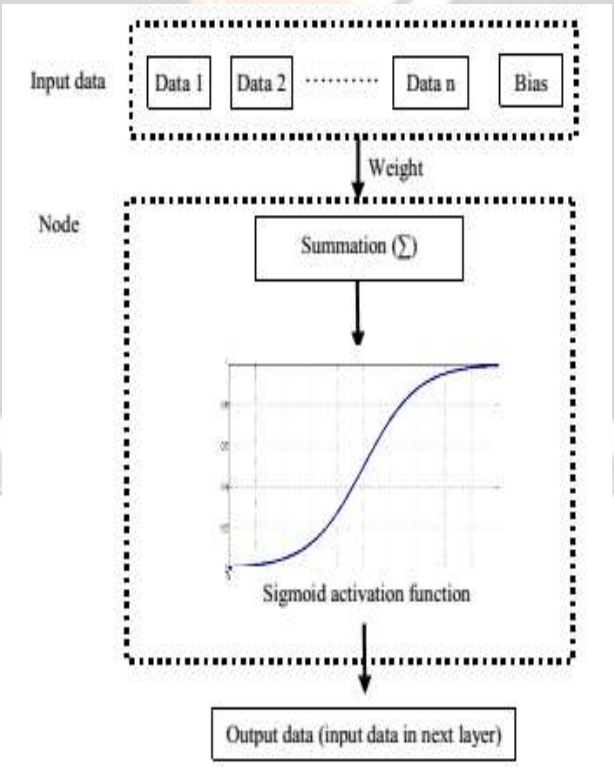


Fig. 1.1.1 Process in one node of neural network [1]

The value of summation (y) is used by the activation function for calculating the new value. The sigmoid activation function is used in all nodes and the new value is input data in the next layer. In each round, data passes all layers.

This process calls feed forward. Fig shows the calculated processes in one node of Neural network. Input data are used summation and sigmoid activation function to create output data for using in the layer.

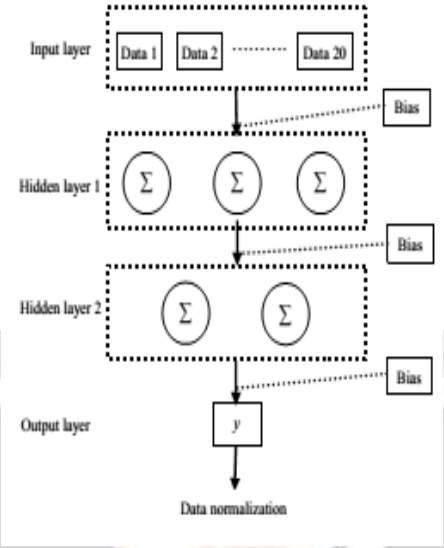


Fig. 1.1.2 Structure of neural network [1]

The next step is back-propagation. This step updates weight in each node from the error. This process training MLP iteratively and finis ing in the next process. hes when the accuracy of the training data set is stable. Finally, the output data equal number of classes are acquired for us.

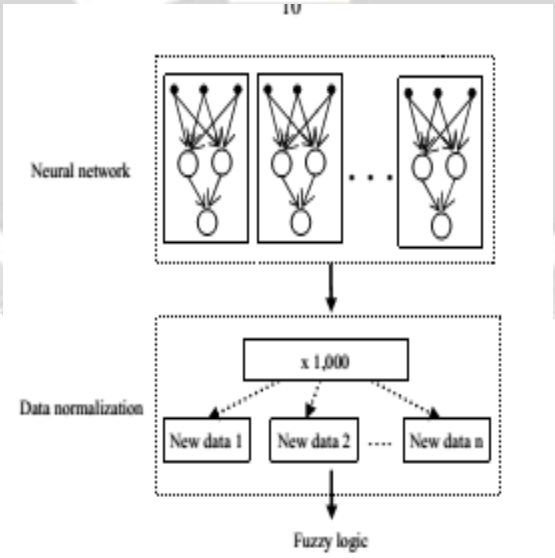


Fig. 1.1.3 Process of Data Normalization [1]

The second process is the data normalization. This process normalizes data from previous process with the decimal technique because the value of data is very small. The process of this step is shown The last sub-process in recognition phase is the Fuzzy logic. This process used the normalized data to create the new feature. First step, all data are used Trapezoidal membership function for creating the membership value and classification rule. Finally,

fuzzy output is used by de fuzzification. The recognition is finished and we have data with compound new feature from fuzzy logic and features from data. Normalization which number of feature equal number of classes. Next process is a classification with K-nearest Neighbors (KNN).

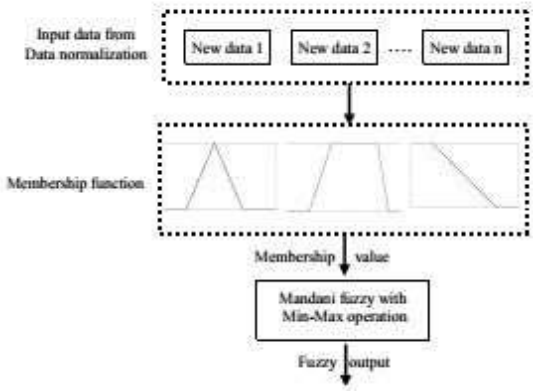
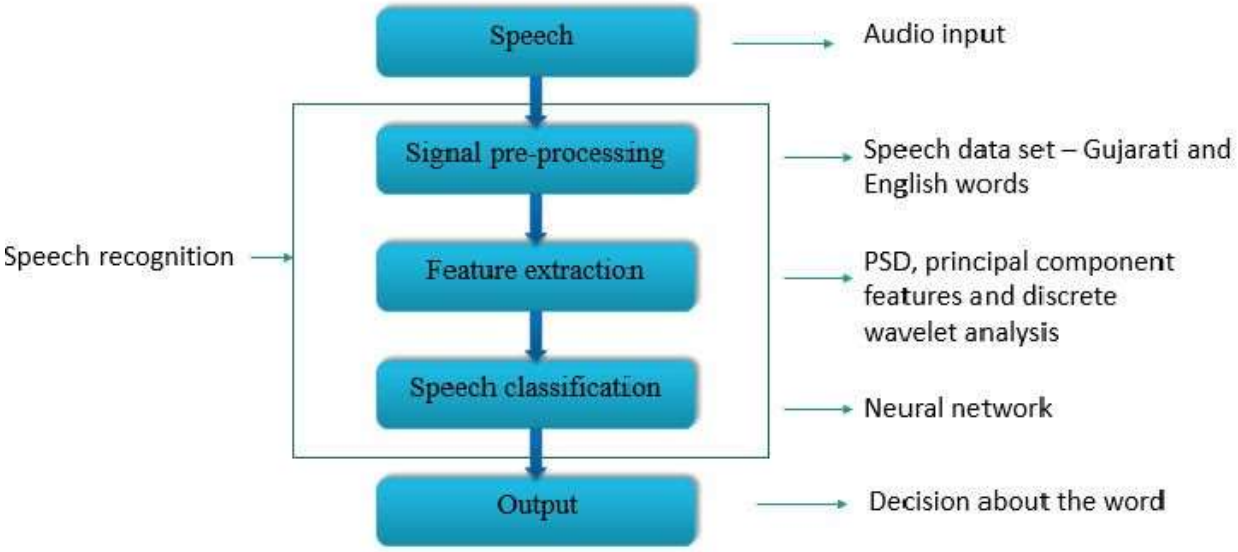


Fig. 1.1.4 Process of Fuzzy Logic [1]

The experiments in this work designed to verify the accuracy and the robustness of the Neuro-fuzzy system. For the speech recognition problems 3 factors are considered, including the number of syllables, different style of pronunciation and environmental with different noises. The experiments are divided into two parts. First part uses data which are cleared noise out for testing accuracy of recognition with different number of syllables and different pronunciations factor. The second part uses data with noise for testing the robustness of the algorithm on noise factor.

2. Block Diagram of Proposed System



The speech signal is given at the input which is to be recognized by the system with the help of algorithm. In this, some English and Gujarati words are given at the input which is then processed with the help of feature extraction. The main features that are used in this system are PSD (power spectral density) , PCA (principal component

analysis) and discrete wavelet analysis. This is then classified with the help of Neuro fuzzy network where the datasets are trained and tested in order to be recognized at the output.

Power Spectral Density (PSD) - It describes the power present in the signal as a function of per unit frequency of the signal. Commonly expressed in watts per hertz where frequency is in terms of Mel frequency.

Principal Component Analysis (PCA) - It is a statistical procedure that uses an orthogonal transformation to convert a set of observations that are possibly correlated.

Discrete Wavelet Transform (DWT) – In this, the wavelets are discretely sampled and has temporal resolution. It captures both frequency and time location information**.**

3. Implementation Flow

- Audio signals are given to the input. (Words through microphone)
  The speech signals are converted from analog signals to digital signals using Fast Fourier transform.
- The signals are computed to obtain the normalized Power Spectrum Density.
- The PSD of the signals are plotted.
- The signals are trained and tested using Neural Network.
- The signals are trained and tested using Neuro-Fuzzy network.
- The text signals are displayed at the output.

4. Conclusion

Many factors relative to human speech are considered in this research in order to test the performance using Neuro fuzzy methodology as a speech recognition algorithm. It show that Neuro fuzzy methodology is robust to many factors and has higher accuracy in recognition and classification then other popular algorithms. It shows that the loud and unpredictable environment noise factor has the highest effect to reduce accuracy in classification.

5. REFERENCES

[1] Srijiranon, K., & Eiamkanitchat, N. (2015, June). Thai speech recognition using Neuro-fuzzy system. In *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2015 12th International Conference on* (pp. 1-6). IEEE.

[2] Yadav, N., Rossato, S., Kahn, J., & Bonastre, J. F. (2016, March). Speaker specific features and phonemes in speech: a proposal for evaluating a possible interaction. In *2016 4th International Conference on Biometrics and Forensics (IWBF)* (pp. 1- 6). IEEE.

[3] Gupta, D., Bansal, P., & Choudhary, K. (2016, January). Noise robust acoustic signal processing using a Hybrid approach for speech recognition. In*Cloud System and Big Data Engineering (Confluence), 2016 6th International Conference* (pp. 489-492). IEEE

[4] Singh, M., & Verma, K. (2011). Speech recognition using neural networks.*International Journal of Technology and Engineering System*, *12*(1), 108- 110.

[5] Helmi, N., & Helmi, B. H. (2008, October). Speech recognition with fuzzy neural network for discrete words. In *2008 Fourth International Conference on Natural Computation* (Vol. 7, pp. 265-269). IEEE.

[6] Samanazari, M., Rajabi, S., Ramezani, A., & Chaibakhsh, A. (2016, January). Fault diagnosis of an industrial once-through benson boiler by utilizing adaptive neuro-fuzzy inference system. In *2016 6th Conference on Thermal Power Plants (CTPP)* (pp. 32-37). IEEE

[7] Sabah, R., & Ainon, R. N. (2009, May). Isolated digit speech recognition in Malay language using neuro-fuzzy approach. In *2009 Third Asia International Conference on Modelling & Simulation* (pp. 336-340). IEEE.

[8] M. Forsberg, "Why is Speech Recognition Difficult?",Department of Computing Science, Chalmers University of Technology, February 24, 2003.

[9] P. P. Bonissone, "Adaptive Neural Fuzzy Inference Systems (ANFIS): Analysis and Applications", technical report, GE CRD, Schenectady, NY USA, 2002.

[10] R. Rabiner and R. W. Schafer, "Digital Speech Processing", Prentice-Hall, 1978.