ADAPTIVE NETWORK BASED FUZZY INFERENCE SYSTEM FOR SPEECH RECOGNITION THROUGH SUBTRACTIVE CLUSTERING

Samiya Silarbi[1], Bendahmane Abderrahmane[2]; and Abdelkader Benyettou[3]

Faculty of mathematical and computer science, department of Computer Science, University of Sciences and Technology Oran USTO-MB, Algeria.

ABSTRACT

Fuzzy modeling require two main steps which are structure identification and parameter optimization, the first one determines the numbers of membership functions and fuzzy if-then rules, while the second identifies a feasible set of parameters under the given structure. However, the increase of input dimension, rule numbers will have an exponential growth and there will cause problem of "rule disaster". In this paper, we have applied adaptive network fuzzy inference system ANFIS for phonemes recognition. The appropriate learning algorithm is performed on TIMIT speech database supervised type, a pre-processing of the acoustic signal and extracting the coefficients MFCCs parameters relevant to the recognition system. First learning of the network structure by subtractive clustering, in order to define an optimal structure and obtain small number of rules, then learning of parameters network by hybrid learning which combine the gradient decent and least square estimation LSE to find a feasible set of antecedents and consequents parameters. The results obtained show the effectiveness of the method in terms of recognition rate and number of fuzzy rules generated.

KEYWORDS

ANFIS, subtractive clustering, Phoneme, recognition.

1. INTRODUCTION

Automatic Speech Recognition has achieved substantial success in the past few decades but more studies are needed because none of the current methods are fast and precise enough to be comparable with human recognition abilities [1,2]. Many algorithms and schemes based on different mathematical paradigms have been proposed in an attempt to improve recognition rates [3,4,5,6].

Neuro-fuzzy modeling is a combination of fuzzy logic and neural network that takes advantage of both approaches, process imprecise or vague data by fuzzy logic [7] and at the same time by introducing learning through neural network. Several architectures have been proposed depending on the type of rule they include Mamdani or Sugeno [8] [9] one of the most influential fuzzy models has been proposed by Robert Jang in [10] called Adaptive Network Based Fuzzy Inference System ANFIS. The rule base of this model contains the fuzzy if-then rule of Takagi and Sugeno's type in which consequent parts are linear functions of inputs instead of fuzzy sets, reducing the number of required fuzzy rules.

The identification of fuzzy model consists of two major phases: structure identification and parameter optimization. The first phase is the determination of number of fuzzy if-then rules and

DOI : 10.5121/ijaia.2014.5604

membership functions of the premise fuzzy sets while the second phase is the tuning of the parameter values of the fuzzy model [11]. However, there will be two problems if we directly use the traditional Takagi Sugeno model for the speech recognition system. The first one is the network reasoning will fail if the input dimension is too large. The second problem is that with the increase of input dimension, rule numbers will have an exponential growth and cause "rule disaster". Thus, determination of an appropriate structure becomes an important issue, then clustering techniques are applied to solve this problem. [12] [13] [14].

This paper present a neural fuzzy system ANFIS for speech recognition. The appropriate learning algorithm is performed on TIMIT speech database supervised type, a pre-processing of the acoustic signal and extracting the coefficients MFCCs parameters relevant to the recognition system. Subtractive clustering is applied in order to define an optimal structure and obtain small number of rules, then learning of parameters network by hybrid learning which combine the gradient decent and least square estimation LSE.

The paper is organized as follows: the section 2 reviews the literature research work, in section 3 details the structure and Principe of learning of ANFIS, while section 4 describes subtractive clustering; experimental results and discussion were detailed in section 5. Section 6 concludes the paper.

2. RELATED WORK

The ANFIS has the advantage of good applicability because it can be interpreted as local linearization modeling, and even as conventional linear techniques for both state estimation and state control which are directly applicable. This adaptive network has good ability and performance in system identification, prediction and control has been applied in many different systems. Since there are not many research works used ANFIS in speech recognition, it is necessary to carry on the exploration and the thorough research on it. ANFIS was used for Speaker verification [15] using combinational features of MFCC; Linear Prediction Coefficients LPC and the first five formants; Recognition of discrete words [16], Speech emotion verification [17] based on MFCC for real time application. In [18] ANFIS was performed to reduce noise and enhance speech, also for the recognition of isolated digits with speaker-independent [19]. Speaker, language and word recognition was completed by ANFIS [20], furthermore for caller behavior classification [21]. All of these works had used clustering techniques to determine the structure of ANFIS. Another work: an automated gender classification is performed by ANFIS [22].

3. ADAPTIVE NETWORK BASED FUZZY INFERENCE SYSTEM ANFIS

3.1. The Concept and Structure

ANFIS proposed by Jang in 1993 multi-layered neural network which connections are not weighted or all weights equal 1**[10]**, is alternate method which combines the advantages of two intelligent approaches neural network and fuzzy logic to allow good reasoning in quantity and quality. A network obtained has an excellent ability of training by means of neural networks and linguistic interpretation of variables via fuzzy logic. The both of them encode the information in parallel and distribute architecture in a numerical framework. ANFIS implement a first order Sugeno style fuzzy system; it applies the rule of TSK Takagi Sugeno and Kang form in its architecture.

Rule: if x is A1 and y is B1 then $f(x) = px + qy + r$

Where x and y are the inputs, A and B are the fuzzy sets, f are the output, p, q and r are the design parameters that determined during the training process. ANFIS is composed of two parts is the first part is the antecedent and the second part is the conclusion, which are connected to each other by rules in network form. Five layers are used to construct this network. Each layer contains several node sits structure shows in figure 1.

layer1: executes a fuzzification process which denotes membership functions (MFs) to each input. In this paper we choose Gaussian functions as membership functions:

$$O_i^1 = \mu_{Ai} = \exp\left(\frac{-(x-c)^2}{\sigma^2}\right) \tag{1}$$

layer2: executes the fuzzy AND of antecedents part of the fuzzy rules

$$o_i^2 = w_i = \mu_{Ai}(x_1) \times \mu_{Bi}(x_2), i = 1,2,3,4 \tag{2}$$

layer3: normalizes the MFs

$$o_i^3 = \overline{w_i} = \frac{w_i}{\sum_{j=1}^{4} w_j}, i = 1,2,3,4 \tag{3}$$

layer4: executes the conclusion part of fuzzy rules

$$O_i^4 = \overline{w_i} y_i = \overline{w_i}(\alpha_1^i x_i + \alpha_2^i x_2 + \alpha_3^i), i = 1,2,3,4. \tag{4}$$

layer5: computes the output of fuzzy system by summing up the outputs of the fourth layer which is the defuzzification process.

$$O_i^5 = overll\_output = \sum_{i=1}^{4} \overline{w_i} y_i = \frac{\sum_{i=1}^{4} w_i y_i}{\sum_{i=1}^{4} w_i} \tag{5}$$

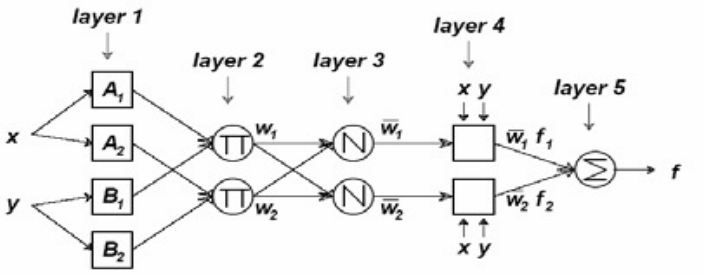

Figure1. ANFIS architecture.

Circles in ANFIS represent fixed nodes that predefined operators to their inputs and no other parameters but the input participate in their calculations. While square are the representative for adaptive nodes that affected by internal parameters.

3.2. Learning Algorithm

The parameters to be tuned in an ANFIS are the membership function parameters of each input, the consequents parameters also number of rules.

$$nbr\_rule = m^n \tag{6}$$

Where n is number of inputs and m number of membership functions by input and they generate all the possible rules.

Two steps of training are necessary which are: Structure learning which allows to determinate the appropriate structure of network, that is, the best partitioning of the input space (number of membership functions for each input, number of rules). And parametric learning carried out to adjust the membership functions and consequents parameters. In most systems the structure is fixed a priori by experts. In our work we combine both of learning sequentially

The subsequent to the development of ANFIS approach, a number of methods have been proposed for learning rules and for obtaining an optimal set of rules. For example, Mascioli et al **[23]** have proposed to merge Min-Max and ANFIS model to obtain neuro-fuzzy network and determine optimal set of fuzzy rules. Jang and Mizutani **[24]** have presented application of Lavenberg-Marquardt method, which is essentially a nonlinear least-squares technique, for learning in ANFIS network. In another paper, Jang **[25]** has presented a scheme for input selection and **[26]** used Kohonen's map to training.

Four methods have been proposed by Jang **[11]** for update the parameters of ANFIS:

- All parameters are update by only gradient decent.
- In first the consequents parameters are obtained by application of least square estimation LSE only once and then the gradient decent update all parameters.
- Sequential LSE that is using extended Kalman filter to update all parameters.
- Hybrid learning: which combine the gradient decent and LSE to find a feasible set of antecedents and consequents parameters.

The most common training algorithm is the hybrid learning. this algorithm is carried out in two steps: forward pass and backward pass, Once all the parameters are initialized, in forward pass, functional signals go forward till fourth layer and the consequents parameters are identified by LSE. After identifying consequents parameters, the functional signals keep going forward until the error measure is calculated. In the backward pass, the error rates propagate backward and the premise parameters are updated by gradient decent.

The function to be minimized is Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\left(\frac{1}{N} \sum_{i=1}^{N} (d_i - o_i)^2 \right)} \tag{7}$$

Where d_i is the desired output and O_i is the ANFIS output for the ith sample from training data and N is the number of training samples.

4. SUBTRACTIVE CLUSTERING

Clustering is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset shares some common features, often proximity according to some defined distance measure. From the machine learning perspective, clustering can be viewed as unsupervised learning of concepts [27][28].

1. Compute the initial potential value for each data point x_i is defined as: $P_i = \sum_{j=1}^{n} \exp\left[- \frac{\left\| x_i - x_j \right\|^2}{\left(ra / 2 \right)^2} \right]$ (8)

2. Where ra is a positive constant representing a neighborhood radius. Therefore, a point would have a height potential value if it has more neighbor points close to itself.

3. the point with the highest potential value is selected as le first cluster center: First cluster center x_{c1} is chosen as the point having the largest density value D_{c1}

4. $P^{(1)*} = \max_i \left(P^{(1)}(x^i) \right)$ (9)

5. The potential value of each data point x_i is reduced as follows: $P_i = P_i - P_{c1} \exp\left[- \frac{\left\| x_i - x_{c1} \right\|^2}{\left(rb / 2 \right)^2} \right]$ (10)

Where ($rb = \beta * ra$) is a positive constant represent the radius of the neighborhood for which significant potential reduction will occur.

β is a parameter called as squash factor, which is multiply by radius values to determine the neighboring clusters within which the existence of other cluster centers are discouraged.

After revising the density function, the next cluster center is selected as the point having the greatest potential value. This process is repeated to generate the cluster centers until maximum potential value in the current iteration is equal or less than the threshold.

5. STRUCTURE LEARNING ALGORITHM FOR T-S FUZZY NEURAL NETWORK

ANFIS offers three approaches to identify cluster namely grid partitioning, subtractive clustering and fuzzy c-means clustering. Grid partitioning approach is useful if the number of features is no more than 6 or 7. If the number of features is too high then this method will cannot be used as the memory requirement will be insufficient when using MATLAB. For fuzzy c-means method, the number of clusters for the dataset needs to be specified. Since no prior knowledge on the number of clusters is available, subtractive clustering will be ideal.

The radius of the clusters will be used as the basis of the cluster formation. A range of radius from 0.0 to 2.2 has been chosen in order to produce optimum number of clusters for the initial fuzzy inference system (FIS). The initial FIS then was fed to the ANFIS for refining the FIS so that it will tune the supervised learning iteratively with more detailed rules generated.

Each cluster center represents a fuzzy if-then-rule. The nth column of cth cluster center is assumed to be the mean value (c_{in}) of the associated Gaussian membership function defined for cth fuzzy set of nth input variable. Then, the standard deviation (a_{in}) of the above mentioned Gaussian functions are calculated as below:

$$a_{in=ra}{}_{n}\left(\frac{\max(x^{n})-\min(x^{n})}{\sqrt{8}}\right) \tag{11}$$

Therefore the cluster centers and squash factors may be viewed as parameters, which the number of fuzzy rules of the initial FIS depend on, before the rule base parameters of the FIS is tuned by ANNs in ANFIS.

6. EXPERIMENTAL AND RESULTS

6.1. TIMIT Database

A reduced subset of TIMIT **[29]** -Phonetic Continuous Speech Corpus (TIMIT – Texas Instruments (TI) and Massachusetts Institute of Technology (MIT)) used in our work, which is the abridged version of the complete testing set, consists of 192 utterances, 8 from each of 24 speakers (2 males and 1 female from each dialect region). In general, phonemes can grouped into categories based on distinctive features that indicate a similarity in articulatory, acoustic and perceptual. The major classes of phonetic TIMIT database are: 6 vowels {/ah/, /aw/, /ax/, /ax-h/, /uh/, /uw/}, 6 fricatives {/dh/, /f/, /sh/, /v/, /z/, /zh/} and 6 plosives {/b/, /d/, /g/, /p/, /q/, /t/}.

6.2. Coding Mel-Frequency Cepstral Coefficients MFCC

Before any calculation, it is necessary to effect some operations to shape the speech signal. The signal is first filtered and then sampled at a given frequency (16 KHZ). Pre-emphasis is carried out to raise the high frequencies. Then, the signal is segmented into frames, each frame has a fixed number N=25ms of speech samples (period in which the speech signal can be considered as stationary).Treating small fragments of signal leads to filtering problems (edge effects). In order to reduce edge effects we use weights windows (Hamming window) these are functions that are applied to set of samples in the window of the original signal.

After signal shaping a discrete Fourier transform DFT particularly fast Fourier transform FFT is applied to pass in the frequency domain and for extracting the spectrum signal. Then, the filtering is performed by multiplying the spectrum obtained by filters (triangular filters). It is possible to employ the output of the filter bank as input to the recognition system. However, other factors derived from the outputs of a bank of filters are more discriminative, more robust and less correlated. It is from the cepstral coefficients derived of the filter bank outputs distributed linearly on the Mel scale, these are the Mel-frequency Cepstral coefficients MFCC Each window provides 12 MFCC coefficients and the corresponding residual energy.

6.3. Results

The corpus that we have used in the classification of phonemes consists of 6 vowels, 6 fricatives and 6 plosives: 31514 instances for learning and 12055 instances for test. Applying subtractive clustering to determinate the initial structure of ANFIS, then Hybrid learning to find a feasible set of antecedents and consequents parameters.

We applied ANFIS on the classification of phonemes by varying the radius of the clustering between [0.0- 2.5], results are summarized in the following tables:

With a radius between [0-1.4] we achieved a recognition rate of 100 % but the number of generated rules was too big, so that exceeds 200 rules which has resulted a large runtime time and especially for real-time application, and this for the three classes of phonemes used.

For vowels we have obtained the best recognition rate of 100 % and 6 rules with a radius of 2.0 and 1.9, we have reduced the number of rules from 6 to 4 with a radius of 2.1 and achieved recognition rate of 83 % (table 1).

For fricatives with a radius of 1.4 we have obtained the best recognition rate that reaches 100 % and 5 fuzzy rules, we could reduce the number of rules to 4 with a recognition rate of 80.66 % (table 2).

For plosives we reached 100 % recognition rate and 5 rules with a radius of 1.4 and 1.5 beyond 1.5 we had 3 rules but reduction of recognition (table 3).

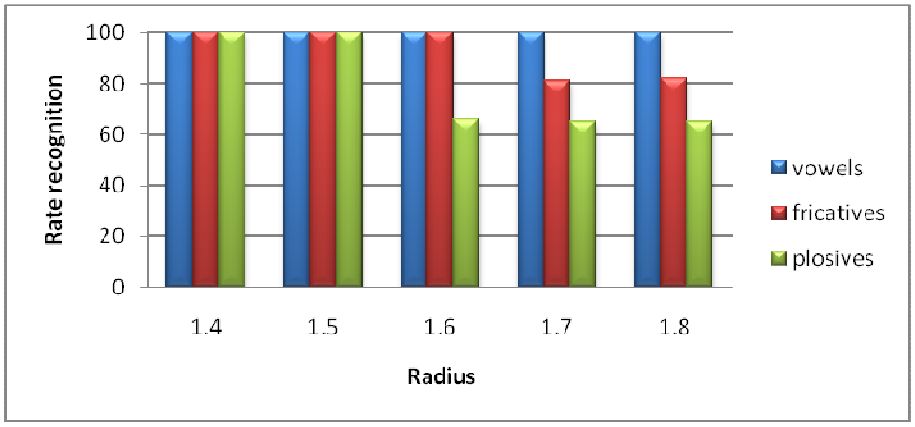

Figure 2. Rate recognition depending of radius

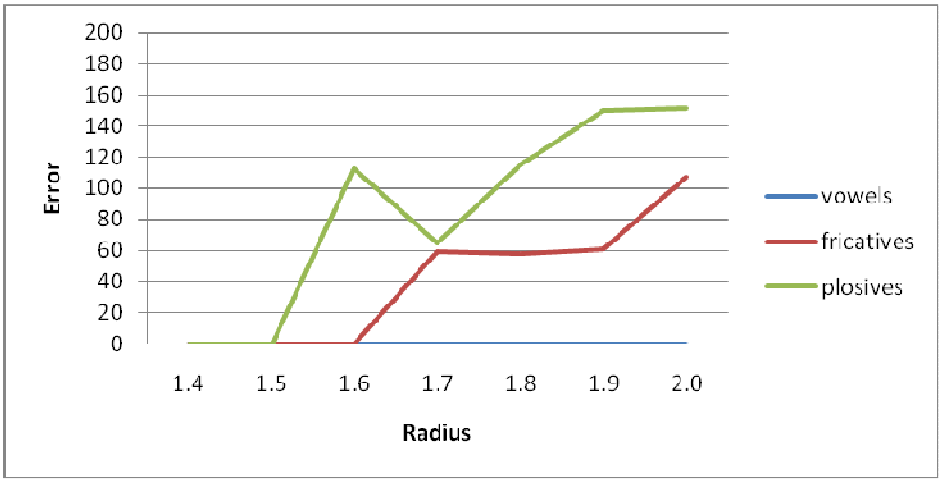

Figure3. Error depending of radius

Table1. Vowels.

Radius	Rate-recognition%		error		Number-rules	
	train	test	train	test	train	test
1.4	100	100	0	0	37	34
1.5	100	100	0	0	22	23
1.6	100	100	0	0	17	15
1.7	100	100	0	0	11	10
1.8	100	100	0	0	9	8
1.9	100	100	0	0	8	6
2.0	100	100	0	0	7	5
2.1	100	83	0	53	7	4
2.2	78	79	66	62	4	4

Table 2. Fricatives.

Radius	Rate-recognition%		error		Number-rules	
	train	test	Train	test	train	test
1.4	100	100	0	0	5	5
1.5	64.33	100	120	0	3	6
1.6	58.66	100	139	0	3	5
1.7	59	80.66	137	59	30	4
1.8	58	81	138	58	3	4

Table 3. Plosives.

Radius	Rate-recognition%		error		Number-rules	
	train	test	train	test	train	test
1.4	100	100	0	0	6	5
1.5	100	100	0	0	8	5
1.6	100	65.33	0	113	5	3
1.7	89	65	34	65	4	3
1.8	89	65	34	115	3	2

A large number of rules are generated by reducing the radius of clustering so most of data are correctly classified but it takes significant time. Contrariwise, by increasing the radius of clustering we had very few rules so many alternatives are not considered by the fuzzy rules, then a lot of data are not correctly classified.

We could see that there was a compromise between the recognition rate and the number of rules generated, with a radius between [1.5-1.8] we were able to achieve perfection on the recognition rate and number of generated rules which engender a very reasonable computation time.

7. CONCLUSION AND FUTURE WORK

In this paper, we have applied adaptive network fuzzy inference system for phonemes recognition. First learning of the network structure by subtractive clustering, in order to define an optimal structure and obtain small number of rules, then learning of parameters network by hybrid learning which combine the gradient decent and least square estimation LSE to find a feasible set of antecedents and consequents parameters.

The appropriate learning algorithm is performed on TIMIT speech database supervised type, a pre-processing of the acoustic signal and extracting the coefficients MFCCs parameters relevant to the recognition system. Finally, hybrid learning combines the gradient decent and least square estimation LSE of parameters network. The results obtained show the effectiveness of the method in terms of recognition rate and number of fuzzy rules generated.

For future work, consider this adaptive network for whole dataset TIMIT, and improving learning algorithm by tuning parameters of ANFIS by means of evolutionary algorithms.

REFERENCES

[1] R. Reddy (2001), Spoken Language Processing: A guide to Theory, Algorithm, And System Development, Prentice-Hall, New Jersey.

[2] X. He, L. Deng (2008), Discriminative Learning for Speech Recognition: Theory and practice, Morgan & Claypool.

[3] Cole, Ron, Hirschman, Lynette, Atlas, Les, Beckman, Mary, Biermann, Alan, Bush, Marcia, et al (1995) The challenge of spoken language systems: Research directions for the nineties. IEEE Transactions on Speech and Audio Processing, vol 3(issue 1).

[4] Scofield, M. C (1991) Neural networks and speech processing. Amsterdam: Kluwer Academic.

[5] Yallop, C. C. (1990) An introduction to phonetics and phonology. Cambridge, MA: Blackwell.

[6] Carla Lopes and Fernando Perdigão. chapter 14 (2011) Phone Recognition on the TIMIT Database. Book Speech Technologies Edited by Ivo Ipsic, ISBN 978-953-307-996-7, 432 pages, Publisher InTech.

[7] George Bojadziev,Maria Bojadziev (1995) Advances in fuzzy systems applications and theory, vol 5 book Fuzzy Sets, Fuzzy Logic, Applications, World Scientific.

[8] Ajith Abraham (2001) "Neuro Fuzzy Systems: State-of-the-Art Modeling", Techniques in Proceedings of 6th International Work-Conference on Artificial and Natural Neural Networks, IWANN 2001 Granada, Spain, June 13–15, Springer Verlag Germany, vol 2084, pp 269-276.

[9] B. Kosko (1991) "Neural Networks and Fuzzy Systems A Dynamic Systems Approach", Prentice Hall.

[10] J.S.R. Jang (1993) "ANFIS: Adaptive Network Based Fuzzy Inference Systems", IEEE Trans, Syst. Man Cybernet. vol 23, No 3, pp. 665-685.

[11] J.S.R. Jang, C.T.Sun and E.Mizutani (1997) Neuro-fuuzy and soft computing: a computational approach to learning and machine intelligence, London: prentice-Hall international, USA.

[12] Agus Priyono, Muhammad Ridwan, Ahmad Jais Alias, Riza Atiq O. K. Rahmat, Azmi Hassan & Mohd. Alauddin Mohd. Ali (2005) "Generation of fuzzy rules with subtractive clustering". Jurnal Teknologi, 43(D). Universiti Teknologi Malaysia, pp. 143–153.

[13] Chuen-Tsai Sun (1994) "Rule-Base Structure Identification in an Adaptive-Network-Based Fuzzy Inference System". IEEE Transaction on Fuzzy Systems, vol. 2, no. 1. pp. 64 – 73.

[14] Ching-Chang Wong and Chia-Chong Chen (1999) "A Hybrid Clustering and Gradient Descent Approach for Fuzzy Modeling". IEEE Transactions on systems, man, and cybernetics—part b: cybernetics, vol. 29, no. 6. pp. 686 – 693.

[15] V .Srihari, R.Karthik and R.Anitha and S.D.Suganthi (2010) "Speaker verification using combinational features and adaptive neuro-fuzzy inference systems", IIMT'10 December 28-30, Allahabad, UP, India, pp. 98-103.

[16] N.Helmi, B.H.Helmi (2008)"Speech recognition with fuzzy neural network for discrete words", ICNC Fourth International Conference on Natural Computation. vol 07, pp. 265 – 269.
[17] N.Kamaruddin, A.Wahab (2008) "Speech emotion verification system (sevs) based on mfcc for real time application", Intelligent Environments, 2008 IET 4th International Conference on, Seattle, pp. 1-7.
[18] Reem Sabah & Raja N. Ainon (2009) "Isolated Digit Speech Recognition in Malay Language using Neuro-Fuzzy Approach", Third Asia International Conference on Modelling & Simulation AMS, , Bali Asia, pp. 336 – 340.
[19] Jasmin Thevaril and H.K.Kwan (2005) "Speech Enhancement using Adaptive Neuro-Fuzzy Filtering" in Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems ISPACS, December 13-16, pp753 – 756.
[20] Bipul Pandey, Alok Ranjan, Rajeev Kumar and Anupam Shukla (2010) "Multilingual Speaker Recognition Using ANFIS". in 2nd International Conference on Signal Processing Systems (ICSPS) vol 3, V3-714 - V3-718.
[21] Pretesh. B. Patel, Tshilidzi Marwala (2011) "Adaptive Neuro Fuzzy Inference System, Neural Network and Support Vector Machine for caller behavior classification", 10th International Conference on Machine Learning and Applications ICMLA, 18-21 December, vol 1, pp 298 - 303.
[22] Sachin Lakra, Juhi Singh and Arun Kumar Singh (2013) "Automated Pitch-Based Gender Recognition using an Adaptive Neuro-Fuzzy Inference System", International Conference on Intelligent Systems and Signal Processing (ISSP), pp 82 – 86.
[23] Gujarat.Manish Kumar, Devendra P. Garg (2004) "Intelligent Learning of Fuzzy Logic Controllers via Neural Network and Genetic Algorithm", Proceedings of 2004 JUSFA Japan – USA Symposium on Flexible Automation. Denver, Colorado, pp. 1-8.
[24] Mascioli, F.M., Varazi, G.M. and Martinelli, G (1997) "Constructive Algorithm for Neuro-Fuzzy Networks", Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, Vol. 1, pp. 459 -464.
[25] Jang, J.-S. R., and Mizutani, E (1996) "Levenberg-Marquardt Method for ANFIS Learning", Biennial Conference of the North American Fuzzy Information Processing Society, pp. 87 -91.
[26] Jang, J.-S.R. (1996) "Input Selection for ANFIS Learning", Proceedings of the Fifth IEEE International Conference on Fuzzy Systems, Vol. 2, pp. 1493 -1499.
[27] J.Han, M.Kamber (2000) "Data Mining: Concepts and Techniques" chapter 8. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers.
[28] V.Estivill-Castro, J.Yang (2000) "A Fast and robust general purpose clustering algorithm". Pacific Rim International Conference on Artificial Intelligence, pp. 208-218.
[29] Zue, V.; Seneff, S. & Glass J (1990) "Speech database development at MIT: TIMIT and beyond", Speech Communication, Vol. 9, No. 4, pp. 351-356.