


Hybrid PSO-ANFIS for Speaker Recognition

Samiya Silarbi, SIMPA Laboratory, University of Sciences and Technology in Oran Mohammed Boudiaf, Bir El Djir, Algeria

Redouane Tlemsani, SIMPA Laboratory, University of Sciences and Technology in Oran Mohammed Boudiaf, Bir El Djir, Algeria

Abderrahmane Bendahmane, SIMPA Laboratory, University of Sciences and Technology in Oran Mohammed Boudiaf, Bir El Djir, Algeria

 <https://orcid.org/0000-0003-0801-9550>

ABSTRACT

This paper introduces an evolutionary approach for training the adaptive network-based fuzzy inference system (ANFIS). The previous works are based on gradient descent (GD); this algorithm converges very slowly and gets stuck down at bad local minima. This study applies one of the swarm intelligent branches, named particle swarm optimization (PSO), where the premise parameters of the rules are optimized by a PSO, and the conclusion part is optimized by least-squares estimation (LSE). The hybrid PSO-ANFIS model is performed for speaker recognition on CHAINS speech dataset. The results obtained by the hybrid model showed an improvement on the accuracy compared to similar ANFIS based on gradient descent optimization.

KEYWORDS

ANFIS, PSO, PSO-ANFIS, Speaker Recognition

INTRODUCTION

Speaker recognition is the process of automatically recognizing a person by his or her voice by using speaker specific information included in speech waves (Huang, Acero, Hon, & Reddy, 2001; He & Deng, 2008). This technology is widely used in various real world situations such as access controls, telephone applications, PC logins, door control systems. Speaker recognition is a field of research that has been extensively studied in recent decades, but none of the proposed approaches is comparable to that of the human mind in terms of speed and accuracy.

Neuro-fuzzy modeling is an alternative method of speaker recognition, which combines the advantages of two intelligent approaches: neural networks and fuzzy logic in order to take advantage of the discriminative power of the former with the ability of reasoning and deduction of the latter. The model can be trained as a neural network with a linguistic interpretation of variables by fuzzy logic (Bojadziev & Bojadziev, 1995). Both encode information at the same time and distribute the architecture in a numerical framework. Several architectures have been proposed, depending on the type of rules they include Mamdani's or Sugeno's (Abraham, 2001; Kosko, 1991). The Adaptive Network-based Fuzzy Inference System (ANFIS) model is one of the most influential fuzzy models proposed by Robert Jang (Jang, 1993) widely used (Fatemeh & Zahra, 2018; Gunasekaran, Varatharajan & Priyan, 2018; RazaviTermeh, Kornejady, Pourghasemi & Keesstra, 2018). The rule base of this model contains fuzzy 'if-then' rules of the Takagi and Sugeno's type in which the conclusion parts are linear functions of the inputs rather than fuzzy sets, thus reducing the number of required fuzzy rules.

DOI: 10.4018/IJCINI.20210401.oa7

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Fuzzy model building involves two important phases: structure identification (which is the determination of number of fuzzy if-then rules and membership functions of the premise fuzzy sets); and, then optimization of these parameters (Jang, Sun, & Mizutani, 1997). Optimization of these parameters is one of the main issues of the ANFIS training, classical learning is based gradient descent, this algorithm has the disadvantage of falling into poor local minima since it is limited to a reduced search space in the neighborhood of an initial random solution that is not always suitable.

In this work, we propose an alternative training approach to optimize the ANFIS parameters more efficiently than the gradient method, the proposed approach is particle swarm optimization (PSO). The PSO algorithm looks for solutions in a larger search space that depends on the number of initial solutions (randomly generated) called swarms. The PSO technique is usually time-consuming to execute, the time required for a good convergence depends on the number of swarms and iterations, which depend on the number of parameters to be optimized. In order to reduce the cost of training, we propose to apply the PSO only to the premise part of the rules and to use a least square estimation (LSE) on the conclusion part. The proposed learning algorithm is tested on the CHAINS data set for speaker recognition. The obtained results are compared with those of the ANFIS trained by the gradient approach.

The rest of the paper is organized as follows: section 2 reviews related literature research on the topic of speaker recognition, section 3 describes the ANFIS model and PSO algorithm, then explain the development of the PSO-ANFIS model for speaker recognition, while section 4 details the experimental results and their discussion, finally Section 5 concludes the paper.

RELATED WORKS

Automatic Speaker Recognition has achieved substantial success in the past few decades. Many works based on different mathematical paradigms have been proposed in order to improve recognition rates. (Tolba, 2011) applied Hidden Markov Model HMM to identify ten Arabic speakers where the (MFCCs) were selected to describe the speech signal. The identification rate was 100% during text dependent experiments and 80% for the text-independent experiments. (Daqrouqa & Tutunjib, 2015) used formants and wavelet packet entropy as inputs to a neural network for classification. Recognition rate was found 90.09% for vowel-dependent and 82.50% for vowel-independent. The Generalized Regression Neural Network GRNN and the Back-Propagation Neural Network BPNN were used as the classifiers (Wu & Tsai, 2011) to classify thirty-six speakers speaking Chinese using empirical mode decomposition EMD feature extraction method. The first method gave a result of 78.16% and the second one gave a result of 88.82%. The artificial neural network NN model is used as classification model (Chen, Wang, Lin, & Li, 2012). MFCCs feature was extracted from speech signal. Then, the dimension of the inputting eigenvector is reduced, and the redundant information is got rid of by means of Kmean Linde-Buzo-Gray algorithm. the error of training network was 0.015 where hidden layer number is 950. (Ge, Iyer, Cheluvaraja, Sundaram & Ganapathiraju, 2017) proposed a neural network NN framework for text-independent speaker classification and verification, using TIMIT 8K database. The 39 (MFCCs) were generated from the preprocessed speech. The system achieved 100% classification. ANN classifier with Back Propagation (BPNN) are used (Wali & Hatture, 2015), the extracted sixteen MFCC features from 50 users, the identification accuracy of around 92% to 70% is achieved for 10 to 50 users.

Multiclass SVM based speaker cluster method have been proposed (Apsingekar & De Leon, 2009) feature vectors composed of 13 MFCCs, 13 delta-MFCCs are appended after CMS and RASTA applied on to MFCCs. Experiments performed on NIST-2002 speech corpus for 64 speaker accuracy was 97%. (Campbell, Campbell, Reynolds, Singer & Torres-Carrasquillo, 2006) applied SVM; GMM; and SVM fused with a GMM for speaker and language recognition on the NIST 2003 data. where 38 MFCCs and 36 LPC were used to describe the speech signal. The equal error rate EER was for SVM 6.46%, for GMM was 7.47% and for SVM/GMM was 5.55%. As well the discriminative

classifier SVM combined with the traditional GMM pattern classification with a cepstral feature vector 39-dimensional MFCC extracted from the speech was proposed (Chakroun, Zouari, Frikha&Hamida, 2015) this combination has been tested on 64 speakers from TIMIT database. They achieved 100% in identification rate with the second set of experiments dealt with only 3 utterances per speaker for the training task. (Campbell, Sturim& Reynolds, 2006) constructed a support vector machine kernel using the GMM supervector. They performed experiments on the 2005 NIST speaker recognition (SRE) corpus where A 19-dimensional MFCC vector is extracted from the pre-emphasized speech signal, the equal error rate (EER) was 5.68% and the minimum decision cost value (minDCF) was 0.0222. (Dhineshkumar, Ganesh & Sasikala, 2016) proposed the GMM/SVM-based automatic speaker identification performances for three acoustic features namely RASTA-MFCC, Gamma tone Frequency Cepstral Coefficients GFCC, and Mean Hilbert Envelope Coefficients MHEC in different noisy conditions. Experiments were conducted on TIMIT phone labelled database corpus. accuracy were for RASTA-MFCC 70.32% for GFCC 68.49% and for MHEC 73.27% under street Noise. (Algabri, Mathkour, Bencherif, Alsulaiman & Mekhtiche, 2017) presented an Arabic speaker recognition system for forensic applications by using GMM-UBM. Thirty-nine MFCCs were used for feature extraction. All experiments were conducted using speeches of 40 speakers of the KSU Speech Database. They achieved using mobile channel recording, at around 97.8% recognition rate with an EER equal to 1.98%. various types of deep features for text-dependent speaker verification are presented (Liu, Qian, Chen, Fu, Zhang & Yu, 2015). These deep features are all implemented in both the GMM-UBM and identity vector framework. The RSR2015 database was chosen to evaluate all the systems. The best system achieves the EER 0.1%. All these works are summarized in Table 1.

The ANFIS has the advantage of good applicability, ability, and performance in system identification, prediction and control. It has been applied in many different systems. Since not many research works have used it in speech recognition. Table 2 presents some of those who do. Thus, it is necessary to carry on the exploration of the use of ANFIS in speech recognition and to more thoroughly research this topic.

DEVELOPMENT OF PSO-ANFIS MODEL

In this section, the ANFIS and PSO are described: then, the development of PSO-ANFIS model for speaker recognition is explained.

Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is a heuristic approach developed by Kennedy and Eberhart (Kennedy & Eberhart, 1995; Nayyar, Le & Nguyen, 2018) as an evolutionary computational method for solving global optimization problems. It was inspired by the metaphor of social behavior observed among insects and animals such as schools of fish and flocks of birds.

PSO is a population-based search method, the population consists of potential solutions, named particles, where each particle moves towards its best previous position and toward the best particle in the whole swarm. These particles are randomly initialized and freely fly across the multi dimensional search space. During flight, each particle updates its own velocity and position based on the best experience of its own and the entire population.

Let $x_i(t)$ denote the position of particle i in the search space at time t assumed to be discrete. The position of the particle is changed by adding a velocity, $v_i(t)$ to the current position (Nayyar, Garg, Gupta, & Khanna, 2018):

$$V_{new} = wV + \varphi_1 (P_{best} - X) + \varphi_2 (G_{best} - X) \quad (1)$$

Table 1. State art of speaker recognition

Authors	Dataset	Features	Classifier	Results (Accuracy %)
(Tolba, 2011)	10 Arabic speaker	MFCC	HMM	100%
(Daqrouqa & Tutunjib, 2015)	Own dataset	five formants and seven entropies	ANN	90,09% Vowel-dependent 82,50% Vowel-independent
(Wu & Tsai, 2011)	laboratory database 36	EMD	GRNN BPNN	78% 89%
(Chen, Wang, Lin, & Li, 2012)	Own dataset	MFCC Kmeanlbg	ANN	0,015 Error
(Ge, Iyer, Cheluvvaraja, Sundaram & Ganapathiraju, 2017)	TIMIT	39 MFCC	ANN	100%
(Wali & Hatture, 2015)		16 MFCC	ANN BPNN	92% 70%
(Apsingekar & De Leon, 2009)	NIST 2002	MFCC CMS RASTA	SVM	97%
(Campbell, Campbell, Reynolds, Singer & Torres-Carrasquillo, 2006)	NIST 2003	38 MFCC 36 LPC	SVM GMM SVM+GMM	6,46% EER 7,47% 5,55%
(Chakroun, Zouari, Frikha & Hamida, 2015)	TIMIT 64 speakers	39 MFCC	SVM+GMM	100
(Campbell, Sturim & Reynolds, 2006)	NIST 2005	19 MFCC	SVM+GMM	5,68% EER
(Dhineshkumar, Ganesh & Sasikala, 2016)	TIMIT	RASTA MFCC GFCC MHEC	GMM+SVM	70,32% 68,49% 73,27%
(Algabri, Mathkour, Bencherif, Alsulaiman & Mekhtiche, 2017)	KSU	39 MFCC	GMM- UBM	97,8% EER 1,98
(Liu, Qian, Chen, Fu, Zhang & Yu, 2015)	RSR2015	Deep feautres	GMM –UBM	0,1% EER

$$X_{new} = X + V_{new} \quad (2)$$

$$\phi_1 = C_1 rand_1 \quad \phi_2 = C_2 rand_2$$

Where the weights are respectively the inertia w ; $rand_1, rand_2 \sim U(0,1)$; the positive acceleration constants C_1 and C_2 . P_{best} is the local best: the best position of this particle has ever visited until now. G_{best} is the neighbors best, assuming that particles are interconnected in a network; each particle can maintains the best position ever found among its neighbors (Milani & Santucci, 2009). In this work, we assume that Neighbors topology is a complete graph including all particles.

PSO is widely used and rapidly developed by researchers; due to its having some advantages over its alternatives. Namely (Kennedy & Eberhart, 1995; Eberhart, & Shi, 2001, Nayyar & Nguyen, 2018):

Table 2. State art of ANFIS in speech

Authors	Features	ANFIS for	Results (Accuracy %)
(Elwakdy, Elsehely, Eltokhy & Elhennawy, 2008)	wavelet transform	Recognition English language speech signals-isolated words	99%
(Srihari, Karthik, Anitha & Suganthi, 2010)	MFCC; LPC and the first five formants	Speaker verification	7.14% EER
(Helmi, N. & Helmi, B., 2008)	LPC	Recognition of discrete words	58%
(Kamaruddin & Wahab, 2008)	MFCC	Speech emotion verification	93% for angry; 89% for sad and 85.3% for happy
(Sabah & Ainon, 2009)	MFCC	Recognize of Malay speech digits	85.24%
(Pandey, Ranjan, Kumar & Shukla, 2010)	LPC, RC, LPCC, LAR, ARCSIN, LSF	Identification of the speaker, language and the words spoken	83.15%
(Silarbi, Bendahmane & Benyettou, 2014)	MFCC	Phonemes recognition	100%

- PSO is an efficient optimization technique for maintaining the diversity of the swarm;
- PSO is an algorithm that is fast and easy to understand and implement (because fewer particles need to be tuned); and;
- PSO requires only primitive mathematical operators; and is computationally inexpensive in terms of both memory requirements and speed.

Adaptive Network-Based Fuzzy Inference System (ANFIS)

ANFIS proposed by Jang in 1993 multi-layered neural network which connections are not weighted or all weights equal 1 (Jang, 1993), ANFIS implement a first order Sugeno style fuzzy system; it applies the rule of TSK Takagi Sugeno and Kang form in its architecture (Takagi & Sugeno, 1985). This rule produces crisp outputs directly, as it uses polynomials as rule consequences (Li & Yang, 2018).

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

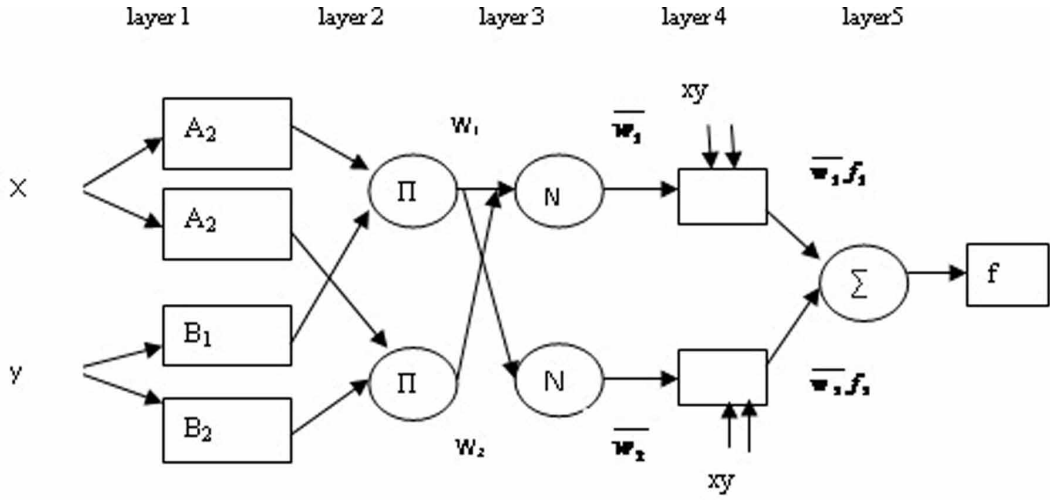
Where x and y are the inputs, A_i and B_i are the fuzzy sets, f are the output, p, q and r are the design parameters that determined during the training process.

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 nodes its structure shows in Figure 1.

Layer 1: executes a fuzzification process which denotes membership functions (MFs) to each input.
In this paper, we choose Gaussian functions as membership functions:

$$O_i^1 = \mu_{A_i} = \exp\left(\left(\frac{-(x-c)^2}{\sigma^2}\right)\right) \quad (3)$$

Figure 1. ANFIS architecture



where c and σ are the center and the standard deviation values of input variable x respectively.

Layer 2: executes the fuzzy AND of antecedents part of the fuzzy rules

$$O_i^2 = w_i = \mu_{A_i}(x_1) \times \mu_{B_i}(x_2), i = 1, 2, 3, 4 \quad (4)$$

Layer 3: normalizes the MFs

$$O_i^3 = \bar{w}_i = \frac{w_i}{\sum_{j=1}^4 w_j}, i = 1, 2, 3, 4 \quad (5)$$

Layer 4: executes the conclusion part of fuzzy rules

$$O_i^4 = \bar{w}_i y_i = \bar{w}_i (\alpha_1^i x_1 + \alpha_2^i x_2 + \alpha_3^i), i = 1, 2, 3, 4. \quad (6)$$

Layer 5: computes the output of fuzzy system by summing up the outputs of the fourth layer which is the defuzzification process.

$$O_i^5 = \sum_{i=1}^4 \bar{w}_i y_i = \frac{\sum_{i=1}^4 w_i y_i}{\sum_{i=1}^4 w_i} \quad (7)$$

Training Algorithm

Two steps of training are necessary. These steps are:

- **Structure Learning:** Allows for determination of the appropriate structure of the network, that is, the best partitioning of the input space (the number of membership functions for each input, and the number of rules). If the input dimension is too large; then rule numbers will have an exponential growth. Thus, determination of an appropriate structure becomes an important issue, so clustering techniques are applied to overcome this problem (Priyono et al., 1999);
- **Parametric Learning:** Carried out to adjust the antecedents and consequents parameters so that a specified objective function is minimized. Jang, Sun & Mizutani (1997) have proposed four methods for updating these parameters, hybrid learning (GD and LSE) is the most common form of training.

This algorithm is carried out in two steps:

- 1) **Forward pass:** The input patterns are propagated to the output and the optimal consequent parameters are estimated by the LSE, while the antecedent parameters are assumed to be fixed in the current cycle through the training set.
- 2) **Backward pass:** The patterns are propagated again and the antecedent parameters are updated by GD, while the consequent parameters remain fixed.

This hybrid procedure is then iterated until the output error is reduced to a desired goal or until a maximum number of training cycles has been reached. This algorithm can become stuck down at local optima. To deal with this problem, evolutionary algorithms can be an ideal solution. These techniques require much computation time if there are many parameters to be optimized. Therefore, the LSE is combined for optimizing a part of the parameters (consequents parameters) and the evolutionary technique is used to optimize the antecedent parameters c_{ij} and σ_{ij} . In this work, we suggest using the PSO technique.

EXPERIMENTAL AND DISCUSSIONS

Database

The CHAINS corpus (Cummins, Leonard, Leonardo & Simko, 2006) contains the recordings of 36 speakers. Of those, 28 are from the Eastern part of Ireland, and speak Eastern Hiberno-English. The rest are from the UK and the USA. Speakers read a variety of texts in a variety of ways: alone; in synchrony with a dialect-matched co speaker; in imitation of a dialect-matched co-speaker; in a whisper; and at a fast rate under a variety of speaking conditions. Four sets of available speakers are employed 8-speaker, 16-speaker, 24-speaker, and 36-speaker sets. In this work we used the 8-speaker and 16-speaker sets. The datasets are in WEKA arff format and contain speech encoded using 25 MFCC.

Initial ANFIS Structure

First all datasets were normalized in the range of (0,1) as follows:

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (8)$$

where X_n and X are the normalized and measured values, respectively. Also, the minimum and maximum values of the measured dataset are X_{\min} and X_{\max} , respectively.

After the determination of the initial ANFIS structure, we use the fuzzy clustering algorithm (FCM) for clustering the data space to determine the number of rules and membership functions for the antecedents and consequents.

For this application, the initial ANFIS structure is generated with 10 rules. Consequently, 10 Gaussian-type membership functions for each input are generated. Each membership function center is initialized with the corresponding fuzzy cluster being obtained by the clustering process with a radius of 0.5.

Training Condition

After the conception of the initial ANFIS, speaker recognition was achieved by ANFIS using both training methodologies:

1. Only ANFIS: Hybrid learning that combines GD and LSE is used, MATLAB reference function (Fuzzy Logic Toolbox, 2000) executes a training routine for Sugeno-type FIS with the following training options: Training epoch number=10; training error goal=0; initial step size= 0.01; step size decrease rate=0.9; and step size increase rate=1.1.
2. Hybrid PSO-ANFIS: An implemented function in the MATLAB environment where the ANFIS antecedent parameters are tuned by a classical PSO.

Hybrid PSO-ANFIS

The optimum PSO parameters can be determined based on several parametric studies. A trial-and-error method was applied in the present paper. The obtained results are summarized in Table 3 and Figure 2.

The function to be minimized is the function for Recognition Rate:

$$Accuracy = \left(\frac{1}{N} \sum (d_i == o_i) \right) \times 100 \quad (9)$$

where d_i is the desired output, O_i is the ANFIS output for the i th sample from the training data, and N is the number of training samples.

Results and Discussions

Figure shows the results obtained by the application of the ANFIS on the 8 speakers of the CHAINS database. The accuracy was 87.99%; while Figure 4 shows the results obtained by the application of the ANFIS on the 16 speakers of the database. The accuracy was 93.80%.

Table 4 shows the results obtained by the PSO-ANFIS model on the 8 speakers by varying the number of particles that is, NPop (size of the population) NPop=20;50;100;150;200;250;300 with 1000 iterations. The experiments carried out have shown that the more we increase the size of the population the more the accuracy likewise increase until reaching 200 particles resulting is an accuracy of 89.17% which is the best result obtained by PSO-ANFIS for 8 speakers. By contrast 300 particles results in diminished accuracy down 88.87%.

Table 5 shows the results obtained by the application of PSO-ANFIS on 16 speakers by varying the size of the population and the number of iterations. The best accuracy obtained for 16 speakers was 94.07% (with number of iterations=1000 and Npop = 100).

Table 6 and Table 7 show a comparison between the two training methods for 8 speakers and 16 speakers, respectively for 8 speakers, there is a clear improvement on the accuracy of 87.99% (ANFIS) at 89.17% (PSO-ANFIS) as well as for 16 speakers of 93.80% (ANFIS) at 94.07% (PSO-ANFIS).

Figure 2. Flowchart of hybrid PSO-ANFIS

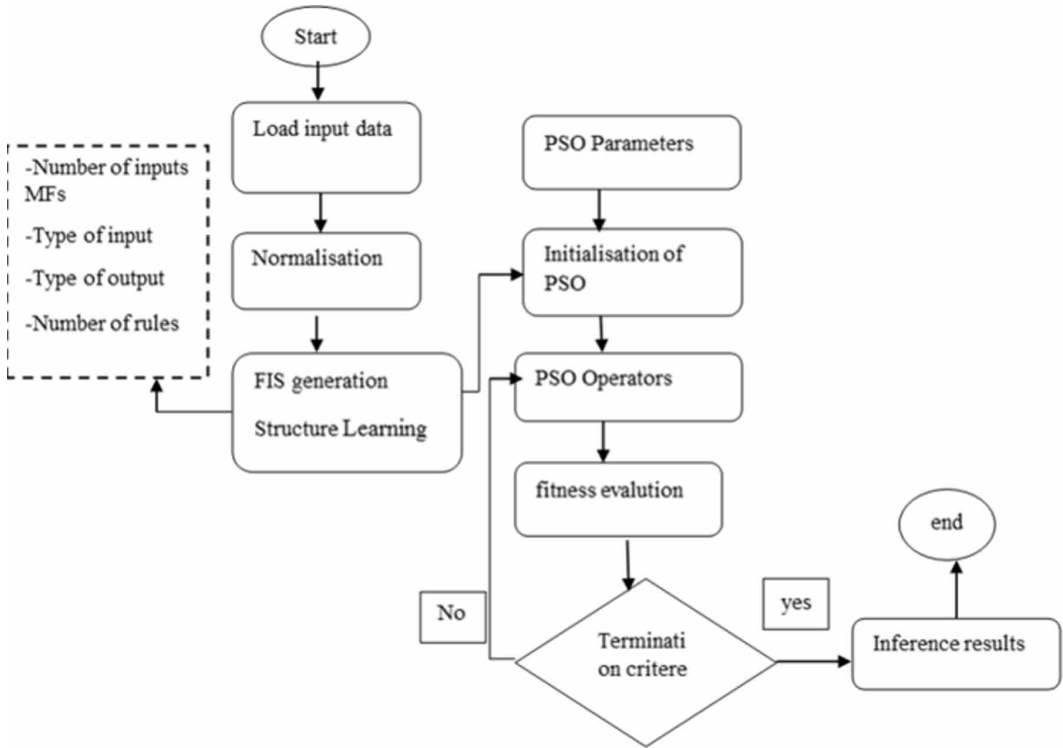


Table 3. PSO parameters

PSO parameter	Value
Inertia Weight W	1
Inertia Weight Damping Ratio wdamp	0.99
Cognitive acceleration C1	1
Social acceleration C2	2
r_1 and r_2	Random

Figure 3. Accuracy on 8 speakers using ANFIS

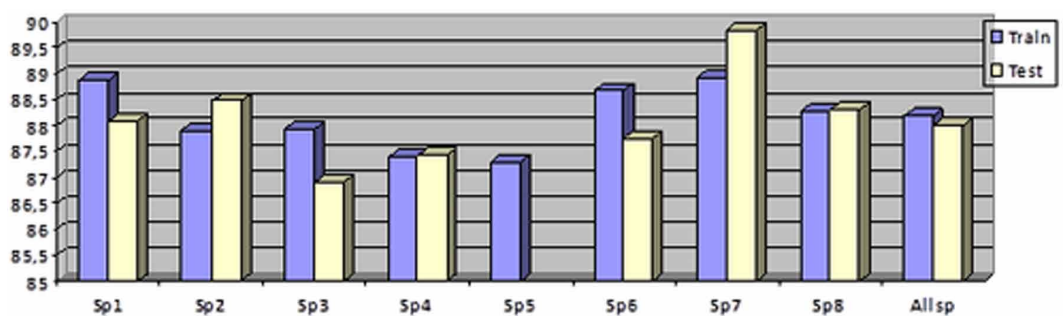


Figure 4. Accuracy on16 speakers using ANFIS

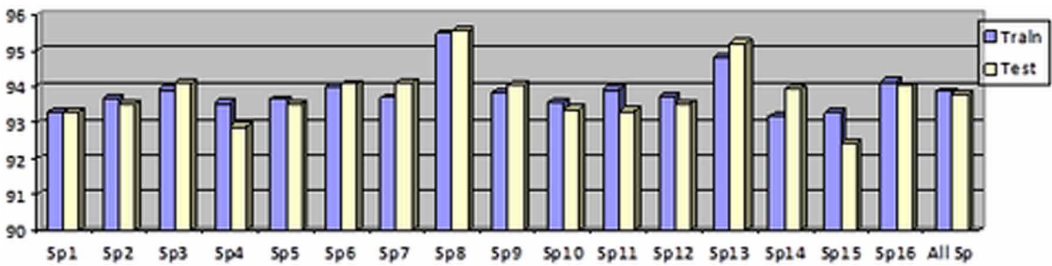


Table 4. Accuracy on 8 speakers using PSO-ANFIS (1000 iterations)

NPop	20	50	100	150	200	300
Train	88.45	89.29	89.35	89.50	89.87	89.50
Test	88.11	88.79	88.81	88.89	89.17	88.87

Table 5. Accuracy on 16 speakers using PSO-ANFIS

Iteration NPop	500 100	500 200	1000 100	1000 200
Train	94.07	94.13	94.21	94.23
Test	93.97	94.01	94.07	94.05

PSO-ANFIS (in comparison of ANFIS alone) offers a clear improvement in accuracy 8 speakers were accurately identified 87.99% of the time by ANFIS but 89.17% of the time by PSO-ANFIS; and 16 speakers were accurately identified 93.80% of the time by ANFIS but 94.07% of the time by PSO-ANFIS. Our work presents a significant improvement comparing to existing works (see Table 8)

Table 6. Comparison of ANFIS and PSO-ANFIS on 8 speakers

Rate	ANFIS	PSO-ANFIS
Train	88.17	89.87
Test	87.99	89.17

Table 7. Comparison of ANFIS and PSO-ANFIS on 16 speakers

Rate	ANFIS	PSO-ANFIS
Train	93.86	94.21
Test	93.80	94.07

Table 8. Comparison of proposed method and existing model

Dataset version	Proposed	(Novakovic, 2011)	(Fazaki& al., 2015)	(Karlos& al., 2016)	(Karlos& al. 2018)
8 Speakers	89.17	MLP 71.77	Self SSL 81.45	SVM 87.44	Entropy 83.94
16 Speakers	94.07	MLP 58.08	Self SSL 78.19	SVM 83.70	SMargin 83.10

CONCLUSION AND PERSPECTIVE

In this work, an Adaptive Network based Fuzzy Inference System (ANFIS) tuned by Particle Swarm Optimization has been developed for speaker recognition. The conventional learning algorithms of ANFIS are based on GD. This algorithm is susceptible to becoming stuck down at local optima during the system training phase hence the exploitation of PSO. The latter was applied to the antecedent part of the ANFIS. Consequent parameters are optimized by LSE, which gives better results as compared to the GD method.

As a perspective, our work will focus on the use of PSO optimization on temporal and recurrent neural models as well as on processing larger benchmarks containing more than 36 speakers

REFERENCES

- Abraham, A. (2001, June). Neuro Fuzzy Systms: State-of-the-Art Modeling Techniques. In Artificial and Natural Neural Networks IWANN, 6th International Work-Conference on (Vol. 2084, pp. 269-276). Academic Press.
- Algabri, M., Mathkour, H., Bencherif, M. A., Alsulaiman, M., & Mekhtiche, M. A. (2017). Automatic Speaker Recognition for Mobile Forensic Applications. *Hindawi Mobile Information Systems*, 2017, 1–6. doi:10.1155/2017/6986391
- Apisingekar, V. R., & De Leon, P. L. (2009, November). Support Vector Machine Based Speaker Identification Systems Using GMM Parameters. In *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, (pp. 1766-1769). IEEE. doi:10.1109/ACSSC.2009.5470201
- Bojadziev, G., & Bojadziev, M. (1995). *Fuzzy Sets, Fuzzy Logic, Applications, Advances in fuzzy systems applications and theory*. World Scientific. doi:10.1142/2867
- Campbell, W. M., Campbell, J. P., Reynolds, D. A., Singer, E., & Torres-Carrasquillo, P. A. (2006). Support vector machines for speaker and language recognition. *Computer Speech & Language*, 20(2-3), 210–229. doi:10.1016/j.csl.2005.06.003
- Campbell, W. M., Sturim, D. E., & Reynolds, D. A. (2006). Support Vector Machines Using GMM Supervectors for Speaker Verification. *IEEE Signal Processing Letters*, 13(5), 308–311. doi:10.1109/LSP.2006.870086
- Chakroun, R., Zouari, L. B., Frikha, M., & Hamida, A. B. (2015, December). A hybrid system based on GMM-SVM for Speaker Identification. In *Intelligent Systems DeSign and Applications(ISDA), 2015 15th International Conference on*, (pp. 645-658). IEEE. doi:10.1109/ISDA.2015.7489195
- Chen, Y., Wang, L., Lin, H., & Li, J. (2012, October). Design of speaker recognition system based on artificial neural network. In *Advanced Optical Manufacturing and Testing Technologies: Optical System Technologies for Manufacturing and Testing. Proceedings of the SPIE, 2012. AOMATT 6th International Symposium on (Vol. 8420, pp. 1-7)*. doi:10.1117/12.970642
- Cummins, F., Leonard, M., Leonardo, T., & Simko, J. (2006, June). The CHAINS corpus CHARacterizing INDividual Speakers. In *Speech and Computer SPECOM, The International Conference on* (pp. 431-435). Academic Press.

Daqrouqa, K., & Tutunjib, T. A. (2015). Speaker identification using vowels features through a combined method of formants, wavelets, and neural network classifiers. *Applied Soft Computing*, 27, 231–239. doi:10.1016/j.asoc.2014.11.016

Dhineshkumar, R., Ganesh, A. B., & Sasikala, S. (2016). Speaker Identification System using Gaussian Mixture Model and Support Vector Machines (GMM-SVM) under Noisy Conditions. *Indian Journal of Science and Technology*, 9(19), 1–6.

Eberhart, R. C., & Shi, Y. (2001). Particle swarm optimization: developments, applications and resources. In *Proceedings of the 2001 Congress on of IEEE international conference on*, (pp. 81-86). IEEE. doi:10.1109/CEC.2001.934374

Elwakdy, A. M., Elsehely, B. E., Eltokhy, C. M., & Elhennawy, D. A. (2008, July). Speech recognition using a wavelet transform to establish fuzzy inference system through subtractive clustering and neural network (ANFIS). In *ICS'08 Proceedings. the 12th WSEAS international conference on Systems* (pp. 381-386). Academic Press.

Fatemeh, Z., & Zahra, Z. (2018). A review of neuro-fuzzy systems based on intelligent control. *Journal of Electrical and Electronic Engineering*, 3(2-1), 58-61.

Fazakis, N., Karlos, S., Kotsiantis, S., & Sgarbas, K. (2015). Speaker Identification Using Semi-supervised Learning. *Lecture Notes in Computer Science*, 9319, 389–396. doi:10.1007/978-3-319-23132-7_48

Fuzzy Logic Toolbox. (2000). Fuzzy Logic Toolbox User's Guide, Version 2. The MathWorks, Inc.

Ge, Z., Iyer, A. N., Cheluvvaraja, S., Sundaram, R., & Ganapathiraju, A. (2017, September). Neural Network Based Speaker Classification and Verification Systems with Enhanced Features. In *Intelligent Systems Conference IntelliSys* (pp. 1–6). IEEE. doi:10.1109/IntelliSys.2017.8324265

Gunasekaran, M., Varatharajan, R., & Priyan, M. K. (2018). Hybrid Recommendation System for Heart Disease Diagnosis based on Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System. *An International Journal Multimedia Tools and Applications*, 77(4), 4379–4399. doi:10.1007/s11042-017-5515-y

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

Helmi, N., & Helmi, B. H. (2008, October). Speech recognition with fuzzy neural network for discrete words. In *Natural Computation 2008, Proceedings ICNC Fourth International Conference on* (Vol. 7, pp. 265–269). IEEE. doi:10.1109/ICNC.2008.666

Huang, X., Acero, A., Hon, H., & Reddy, R. (2001). *Spoken Language Processing: A guide to Theory, Algorithm, And System Development*. Prentice-Hall.

Jang, J. S. R. (1993). ANFIS: Adaptive Network Based Fuzzy Inference Systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 665–685. doi:10.1109/21.256541

Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing: a Computational Approach to Learning and Machine Intelligence*. Prentice-Hall.

Kamaruddin, N., & Wahab, A. (2008, July). Speech Emotion Verification System (SEVS) based on MFCC for real time application. In *Intelligent Environments, IET 4th International Conference on* (pp. 1-7). IEEE.

Karlos, S., Fazakis, N., Karanikola, K., Kotsiantis, S., & Sgarbas, K. (2016). Speech Recognition Combining MFCCs and Image Features. *Lecture Notes in Computer Science*, 9811, 651–658. doi:10.1007/978-3-319-43958-7_79

Karlos, S., Kaleris, K., Fazakis, N., Kanas, V. G., & Kotsiantis, S. (2018). Optimized Active Learning Strategy for Audiovisual Speaker Recognition. *Lecture Notes in Computer Science*, 11096, 281–290. doi:10.1007/978-3-319-99579-3_30

Kennedy, J., & Eberhart, R. C. (1995). Particle swarm optimization. In *Neural Networks, 1995. Proceedings. IEEE International Conference on* (pp. 1942-1948). IEEE. doi:10.1109/ICNN.1995.488968

Kosko, B. (1991). *Neural Networks and Fuzzy Systems A Dynamic Systems Approach*. Prentice-Hall.

- Li, J., Yang, L., Qu, Y., & Sexton, G. (2018). An extended Takagi–Sugeno–Kang inference system (TSK+) with fuzzy interpolation and its rule base generation. *Soft Computing*, 22(10), 3155–3170. doi:10.1007/s00500-017-2925-8
- Liu, Y., Qian, Y., Chen, N., Fu, T., Zhang, Y., & Yu, K. (2015). Deep feature for text-dependent speaker verification. *Speech Communication*, 73, 1–13. doi:10.1016/j.specom.2015.07.003
- Milani, A., & Santucci, V. (2009). Online PSO for Web Marketing Optimization. *IEEE International Conference on e-Business Engineering, Macau, 2009*, 583–587. doi:10.1109/ICEBE.2009.92
- Nayyar, A., Garg, S., Gupta, D., & Khanna, A. (2018). Evolutionary computation: theory and algorithms. In *Advances in Swarm Intelligence for Optimizing Problems in Computer Science* (pp. 1–26). Chapman and Hall/CRC. doi:10.1201/9780429445927-1
- Nayyar, A., Le, D. N., & Nguyen, N. G. (Eds.). (2018). *Advance in Swarm Intelligence for Optimizing Problems in Computer Science*. CRC Press. doi:10.1201/9780429445927
- Nayyar, A., & Nguyen, N. G. (2018). Introduction to Swarm Intelligence. In *Advances in Swarm Intelligence for Optimizing Problems in Computer Science* (pp. 53–78). Chapman and Hall/CRC. doi:10.1201/9780429445927-3
- Novakovic, J. (2011). Speaker identification in smart environments with multilayer perceptron. *2011 19th Telecommunications Forum (TELFOR) Proceedings of Papers*, 1418–1421.
- Pandey, B., Ranjan, A., Kumar, R., & Shukla, A. (2010, July). Multilingual Speaker Recognition Using ANFIS. *Signal Processing Systems (ICSPS), 2010 2nd International Conference on*, 3, 714–718.
- Priyono, A., Ridwan, M., Alias, A. J., Rahmat, R. A. O. K., Hassan, A., & Ali, M. A. M. (2005). Generation of fuzzy rules with subtractive clustering. *Journal Teknologi*, 43(1), 143–153. doi:10.11113/jt.v43.782
- Sabah, R., & Ainon, R. N. (2009, May). Isolated Digit Speech Recognition in Malay Language using Neuro-Fuzzy Approach. In *Modelling & Simulation, 2009. AMS '09. Third Asia International Conference on*, (pp. 336–340). IEEE.
- Silarbi, S., Bendahmane, A., & Benyettou, A. (2014). Adaptive Network Based Fuzzy Inference System For Speech Recognition Through Subtractive Clustering. *International Journal of Artificial Intelligence & Applications*, 5(6), 43–52. doi:10.5121/ijaiia.2014.5604
- Srihari, V., Karthik, R., Anitha, R., & Suganthi, S. D. (2010, December). Speaker verification using combinational features and adaptive neuro-fuzzy inference systems. In *Intelligent Interactive Technologies and Multimedia. IIMT'10 the First International Conference on* (pp. 98–103). doi:10.1145/1963564.1963580
- Sun, C. T. (1994). Rule-Base Structure Identification in an Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Fuzzy Systems*, 2(1), 64–73. doi:10.1109/91.273127
- Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *Syst Man and Cybern IEEE Trans SMC*, 15(1), 116–132. doi:10.1109/TSMC.1985.6313399
- Tolba, H. (2011). A high-performance text-independent speaker identification of Arabic speakers using a CHMM-based approach. *Alexandria Engineering Journal*, 50(1), 43–47. doi:10.1016/j.aej.2011.01.007
- Wali, S. S., & Hatture, S. M. (2015). MFCC Based Text-Dependent Speaker Identification Using BPNN. *International Journal of Signal Processing Systems*, 3(1), 30–34.
- Wong, C. C., & Chen, C. C. (1999). A Hybrid Clustering and Gradient Descent Approach for Fuzzy Modeling. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, 29(6), 686–693. doi:10.1109/3477.809024 PMID:18252349
- Wu, J. D., & Tsai, Y. J. (2011). Speaker identification system using empirical mode decomposition and an artificial neural network. *Expert Systems with Applications*, 38(5), 6112–6117. doi:10.1016/j.eswa.2010.11.013

Samiya Silarbi received her Master degree from University of Sciences and Technology of Oran at 2009 specialty in computer pattern recognition and artificial intelligence, supporting the theme "recognition of individuals by the geometry of the hand", and prepare a Ph.D in speech recognition and neuro-fuzzy logic in SIMPA" (Signal Image Parole) Laboratory" at USTO-MB University . her current research interest is in automatic speech recognition; neural networks, machine learning, neuro-fuzzy logic, evolutionary algorithms and optimization.

Redouane Tlemsani received their HDR and PhD in Computer Sciences, young researcher in pattern recognition and artificial intelligence and Master in data processing and computer sciences from the university of Sciences and Technologies of Oran USTO, Algeria. They are now an associate professor at University of Sciences and Technologies of Oran Mohamed Boudiaf USTO MB, Algeria. They are head of a research team for LaRATIC laboratory since 2013. They teach courses at both Undergraduate and graduate levels in computer science and software engineering.

Abderrahmane Bendahmane received his Ph.D degree (2018) in Artificial Intelligence from University of Sciences and Technology of Oran Mohamed Boudiaf (USTO-MB). Currently, he is a Professor in Computer Science at USTO-MB and member of Signal Image Parole laboratory where he is conducting research activities in the areas of evolutionary optimisation and machine learning.