In Proceedings of the Sixth International Conference on Cognitive Modeling, 364-365.
Mahwah, NJ: Lawrence Earlbaum.

New Contribution To Adaptive Temporal Radial Basis Function Applied on TIMIT Corpus

Mesbahi Larbi (mesbahi_99@yahoo.com)
SIMPA laboratory, BP 1505, Department of Computer Science,
USTO-MB, Oran, 31000, Algeria.

Benyettou Abdelkader (aek.benyettou@univ-usto.dz)
SIMPA laboratory, BP 1505, Department of Computer Science,
USTO-MB, Oran, 31000, Algeria.

Introduction

A successful speech recognition system has to determine features not only present in the input pattern at one point in time, but also features of input pattern changing over time (e.g., Berthold, 1994; Benyettou, 1995). In network design, great importance must be attributed to correct choice of the number of hidden neurons, which helps avoiding problems of overfitting and contributes to reduce the time required for the training without significantly affecting the network performances (e.g., Colla & Reyneri & Sgarbi, 1999), but never looking to architecture adapting effect according to input.

The goal to combine the approach of the RBF with the shift invariance features of the TDNN, can be get a new robust model, this is named temporal radial basis function "TRBF" (e.g., Mesbahi & Benyettou, 2003), but to be more efficient, we have adapt these networks so that they come more dynamic according to their behaviour and features of the object has study. It can be goes more clearly in continuous speech.

Therefore in object to obtain an Adaptive TRBF, we must adapt the TRBF networks, consequently it was necessary to develop an algorithm that permits to solve this type of problem, this algorithm is called "DOLS" which means Dynamic Orthogonal Least Square, that will be presented in this paper.

Dynamic OLS Algorithm

The OLS with its classic version can not adapt our ATRBF network, therefore the original idea of the Dynamic OLS method resides at each iteration, the creation of a hidden centre block and not only one centre, as the size of every block is expressed like suit :

If we consider that the input vector is composed of n characteristic on a temporal input window of Nfe length and if we take the value of input time delay equal to Nde, such as : Nfe $\geq$Nde. Then the number of neurons composing every block is equal to : Nfe-Nde+1.

The size of every hidden centre is equal to : n x Nde, (Figure 1).

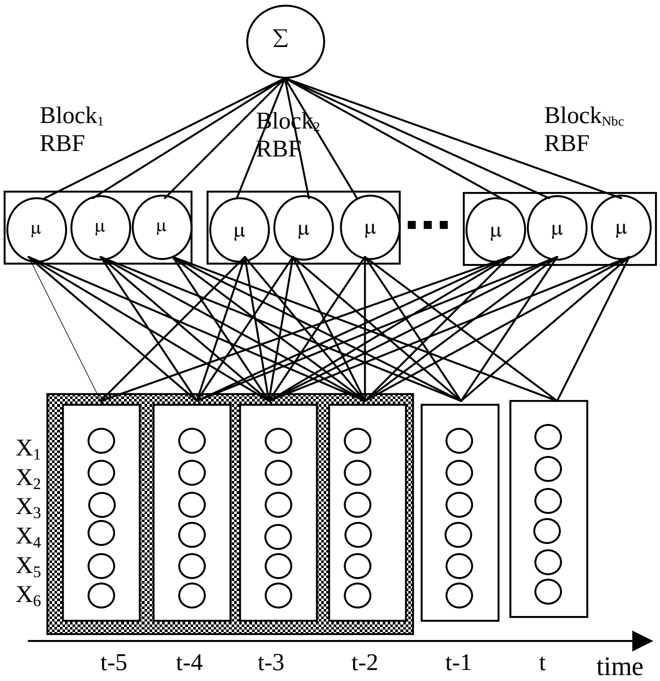


Figure 1: Parameters of this ATRBF network are : n : number of features equal to dim(X)=6, Nde : Time delay equal to 4, Nfe : Size of the input window equal to 6, Nnc : number of hidden neurons by block equal to 3, Nbc: represents the number of centres in hidden bloc, μ : represents a centre of size equal to Nde x n.

With this representation, we can follow the following steps:

$$d = (P\ \theta + E) \qquad (1)$$

The Orthogonalisation of the columns Pi can be gotten by the decomposition of the P matrix in two matrices W and A as:

$$P = W A \qquad (2)$$

Where W : of size Nex M, is the orthogonal image of the P matrix. A : of size M x M, is a superior triangular matrix containing orthogonal coefficients. The A matrix is defined:

Table 1: Representation of A Matrix containing the orthogonal coefficients.

1	$\alpha_{1,2}$				α_{1M}
0	1				$\alpha_{2,M}$
0	0	1			
				$\alpha_{M-2,M-1}$	$\alpha_{M-2,M}$
0			0	1	$\alpha_{M-1,M}$
0			0	0	1

The space begotten by the vectors P_i is the same space begotten by the vector W, then:

$$d = W*G + E \quad (3)$$

Where G=A. θ is the search solution.

$$[err]_i = G_i^2 . w_i^T . w_i /(d^T. d) ; 1 \le i \le M \quad (4)$$

At each iteration we calculate the elements of A and W :

$$\alpha_{jk}{}^i = W^t_j * p_i / W^t_j * W_j \quad (5)$$

$$W^i_k = p_i - \sum^{k-1} \alpha_{jk}{}^i * W_j \quad (6)$$

The criterion of iteration stop here is not based merely on the Akaike criterion:

$$1 - \sum_{i=1}^{i=M} err_i \le \varepsilon \quad (7)$$

Modulo M on Nnc=0, where $M \neq 0$ (8)

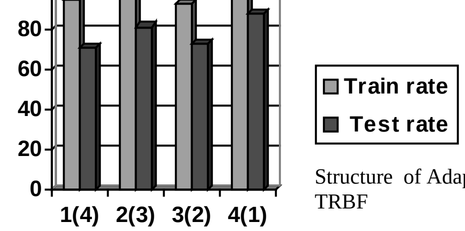
In end of iterations, we calculate the synaptic weights according to the system:

$$G = A * \theta \quad (9)$$

Application

The developed approach has been achieved on a subset of the TIMIT corpus (e.g., Benyettou, 1995; Yak et al, 1995), which is organized of 6 vowels, 6 fricatives and 6 plosives. For our application we reduced the space of study for the case of plosives. Signals have been sampled to 16 KHZS with an analysis Cepstral under the Mel ladder, takes all 20ms in Hamming windows of 25ms giving each 12 MFCCS coefficients and the corresponding residual energy.

This work was achieved on a PIV microcomputer 1.7 GHz with 256 Mo of RAM, developed by the C++ builder and Matlab 6.5 languages. Concerning the training data basis, it has been quantified with a SOM-Kohonen of size 15x15 , generating a training basis of 225 phoneme examples, therefore the size of training basis is about 675 examples, every example is normalized on an input length window equal to 4. The size of the test basis is 450 example at average 150 examples by phoneme. The obtained results can be viewed on the Figure 2



Structure of Adaptive TRBF

Figure 2: Performance according to size of windows x(y), x :Nde, y :Nnc.

The global rate accuracy is about 98% in learning and 83% in test, the Figure 2 shows us the influence of the time delay change on the performance of the network. If the number of the time delay is narrow the performance degrades seen that it has less space time to browse all characteristic.

On the other hand if the time delay is large, there is a risk that the system becomes not shift invariant over time. The best case is choosing the time delay in the median

Conclusion

In this paper we have presented a new approach based on dynamic orthogonal least square developed for adaptive temporal radial basis function , applied to speech recognition. The main advantage with regard to other neural architectures, is the time won in the training, in addition we have few parameters to adjust. The Adaptive TRBF combines advantages of the TDNN that are shift invariant in the time and their capacity of temporal feature recognition and advantages of the RBF in their speed.

We have obtained a 98% in learning rate and 83% in test rate. This shown that our algorithm has good rate of training and test. We hope that this work can be generalized and tested in different domains like pattern recognition and more especially in applications covering the Spatio-Temporal approaches, like mobile robotics, detection of targets, economic fluctuations etc.

Acknowledgments

All my thanks to my supervisor Professor A. Benyettou without missing my Professor N. E. Berrached in master studies director of LARESI Laboratory, Electronic Department, USTO, Oran, Algeria.

References

Berthold, M. R. (1994). A Time Delay Radial Basis Function for Phoneme Recognition. Proceeding of the International Conference on Neural Network, Orlando.

Benyettou, A. (1995). Acoustic Phonetic Recognition in the Arabex System. Int. Work Shop on Robot and Human Communication, ATIP95.44, Japan.

Colla, V., & Reyneri, L.M., & Sgarbi, M. (1999). Orthogonal Least Square Algorithm Applied to the Initialisation of Multi-Layer Perceptron. ESSAN'1999, (pp. 363-369).

Mesbahi, L., & Benyettou, A. (2003). A New Contribution Towards a Temporal Radial Basis Function Applied To Mackey-Glass Time Series. 3 rd ICNNAI Nov 2003, (pp. 72-78). Minsk, Belarus.

Yak, D., & Lin, Q., & che, C., & Jin, L., & Flanagan, J. (1995). Environment-Independent Continuous Speech Recognition. IEEE Automatic Speech Recognition Workshops, Dec 10-13, 1995, (pp.151-152). Utah.