INTELLIGENT APPROACH TO TRAIN WAVELET NETWORKS FOR RECOGNITION SYSTEM OF ARABIC WORDS

Ridha Ejbali, Mourad Zaied
Research Group on Intelligent Machines, University of Sfax, Soukra street, B.P. 1173, 3038, Sfax, Tunisia

Chokri Ben Amar
Research Group on Intelligent Machines, University of Sfax, Soukra street, B.P. 1173, 3038, Sfax, Tunisia

Keywords: Speech recognition, Recognition of Arabic words, Wavelet network, Intelligent algorithm, Randomly training of wavelet network, Intelligent algorithm o training of wavelet network.

Abstract: In this work, we carried out a research on speech recognition system particularly recognition system of Arabic words based on wavelet network. Our approach of speech recognition is divided into three parts: parameterization phase, training phase and recognition phase. This paper aims at introducing an intelligent algorithm of training of wavelet network for words recognition system. It presents also experimental results and a comparison between old training algorithm based on randomly training of wavelet network and our new approach based on intelligent algorithm of training of wavelet network for recognition system of Arabic words.

1 INTRODUCTION

Speech recognition is an electronic vision of human communication. The idea is to interact within heterogeneous worlds, human beings and the electronic one. Despite the simplicity of the idea, it illustrates various difficulties. These problems are of different types, those due to the speech signals representation, and those due to the methods and algorithms adopted in speech recognition.

Having seen the complexity that is becoming attached to the development of speech recognition systems, the specialists proceeded to decomposition into sub-problems. Recognizing speech content means recognizing the units that constitute this speech, for example words, diphthongs or phonemes of each sentence.

To remedy these problems, recognition of speech units, a solution based on the wavelet network (neural network with wavelets as transfer functions) is proposed. The wavelets are an excellent approximators and signal analyzers. Their time-frequency analysis makes them an effective and innovative tool.

In this paper, we introduce a new algorithm to train the wavelet network, which permit an intelligent and effective modelling of acoustic units of training base.

2 THEORETICAL BACKGROUND

The field of wavelet networks is recent, although some attempts have previously been held to build a theoretical basis and several applications in various fields. The use of wavelet networks began with the practice of Gabor wavelets in classification and recognition of images.

The origin of wavelet networks can be traced back to the work of Daugman (Daugman, 2003), in which Gabor wavelets have been used for image classification. These networks became popular after the work of Pati, Zhang and Szu (Pati and Krishnaprasad, 1993), (Zhang and Benveniste, 1992), (Szu, Telfer and Kadambe, 1992). They were introduced as a special feed forward neural network.

Wavelet network allows the representation of a non linear function by training while comparing their inputs and their outputs. This training is made while representing a non linear function by a

Ejbali R., Zaied M. and Ben Amar C..
INTELLIGENT APPROACH TO TRAIN WAVELET NETWORKS FOR RECOGNITION SYSTEM OF ARABIC WORDS.
DOI: 10.5220/0003118905180522
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (KDIR-2010), pages 518-522
ISBN: 978-989-8425-28-7

combination of activation functions. The sigmoid function is often used as an activation one. The input of this prototype is a set of parameters $(t_i)_{1 \le i \le n}$. So the entries are not actual data but only values describing specific positions of the analyzed signal. The hidden layer contains a set of nodes; each node is composed of a translated and dilated wavelet. The output layer contains one node which sums the outputs of the hidden layer by weighted connections weights w_i . Figure 1 illustrates the general form of a wavelet network.

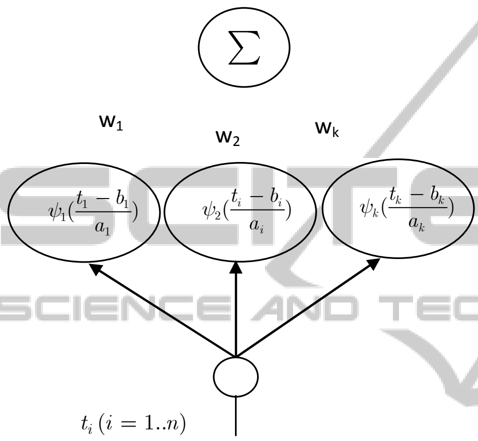

Figure 1: Architecture of wavelet network.

3 PROPOSED APPROACH

3.1 General Approach

The idea of this new architecture, named speech recognition system of Arabic Words based on wavelet networks, is derived from the speech recognition system architecture (Ejbali, Benayed and Alimi, 2009). It is composed of:

- Parameterization module: This subsystem will allow the transformation of all training speech signals to vector characteristics allowing more precise specification. To decode the speech, the recognition system will precede to window each speech signal to extract the acoustic vectors.
- Training module: This module prepares the training models to the recognition system. During the preparation of the training networks, the selection of the wavelet ψ_{i+1} is made randomly. The training process does not verify if the error between the reconstructed vector and the original one, in every stage, decreases or increases but it wait that it reach an error minimal fixed to come to an end.
  This database of training model will include the weights $(w_{i\,(i=1..n, n \in IN)})$ and the wavelets $(\psi_{i\,(i=1..n, n \in IN)})$ of each acoustic unit.
  The learning algorithm described below is too slow and inefficient in the selection of wavelets for the construction of the learning network.
- Recognition module: This module allows the recognition of the speech input. It's based partly on the training basis and on the characteristic vector of input speech. It calculates the new weights of wavelet networks for each network and then evaluates a distance between the new and the old weight. At the end of this phase, a comparison between the distances of the weights will decide on the text given in the input sequence of the system.

Decision is based on distance between weights of networks of recognition phase and each network of training phase. This distance is defined as follows (Zaied, Jemai and Ben Amar, 2008):

(ψ, v) and (ψ, w) are two networks, the distance between them is defined by:

$$D = \left\| \sum_{i=1}^{n} v_i \psi_i - \sum_{j=1}^{n} w_j \psi_j \right\|_2 \quad (1)$$

3.2 Proposed Algorithm of Training

The Training process is preceded by the preparation of the library of candidate wavelets. This library constitutes the base of selection of the activation functions of the wavelet network. After preparation of the library, the algorithm will test and pick up the mother wavelet that covers the support of the signal to be analyzed. The adopted library will be built from the selected wavelets and its translated and dilated form. The first activation function is the lowest frequency wavelet of the library (the mother wavelet). It will be used as a first activation function (in this stage only one neuron is in the hidden layer).

The training is an incremental process. Each time we select the next wavelet of the library (the selection is sequential) and iterate the following steps. The stop of the training process is controlled by an error E_{min} between the input and the output network or a predefined number of wavelet used for the training or a predefined number of neurons in the hidden layer of the network. If the added wavelet ψ_n creates a basis (orthogonal or bi-orthogonal) with the (n-1) activation wavelets of the network, it is used as an activation function of a new neuron in the hidden layer; else it will update the (n-1) old weights of the network. After the construction of the network, the training algorithm calculates the dual basis formed by the activation wavelets of the hidden layer of the network and the new selected wavelet.

Knowing the hidden layer, wavelets and the connection weights, we can calculate the output of the network. If the error E_{min} or the number of wavelets used or the number of neurons are reached then it is the end of training else, another wavelet is selected from the library.

The following figure illustrates the training process:

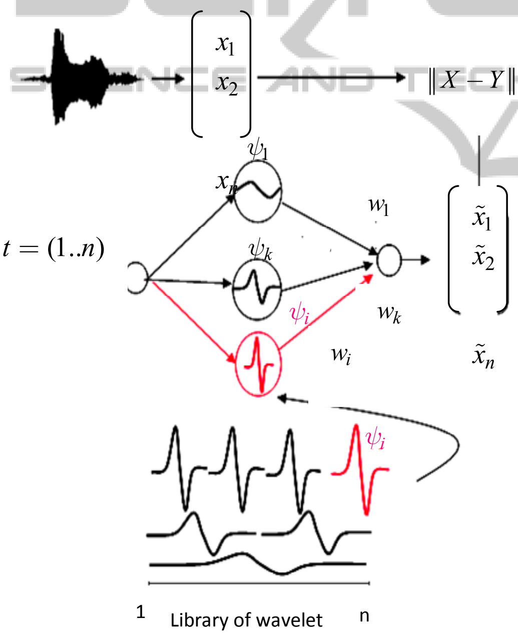

Figure 2: Intelligent training process.

The library of wavelets is constructed by a sampling on a dyadic grid of continuous wavelet transform. This sampling gives in the first scale one wavelet (the mother wavelet). Every time that we climb a scale, the number of wavelet in this scale is multiplied by two.

3.3 Calculation of Weights

To calculate connection weights of a wavelet network, it is necessary to use the same wavelet family (translated dilated) as activation function of each node of the hidden layer (Zhang and Benveniste, 1992).

The calculation of connection weights in each step is possible by projecting the signal to be analyzed on the same family of wavelets: $w_i = < f, \psi_i >$. To use this equation, we need a family of orthogonal wavelets.

To calculate connection weights of networks of recognition, we will use a family of dual wavelet network. Two families of wavelet ψ_i and $\tilde{\psi}_j$ are called biorthogonal if for all i and j we have: $< \psi_i, \tilde{\psi}_i >= \delta_{i,j}$. Wavelet ψ is known as the primal wavelet and wavelet $\tilde{\psi}$ is the dual one. If $\psi_i = \tilde{\psi}_i$, the family of wavelet ψ composes an orthogonal base.

The use of biorthogonal wavelets allows direct calculation of the WN connection weights. Suppose that f is a signal, $(\psi_i)_{1<i\leq n}$ is an orthogonal wavelet family and $(\tilde{\psi}_i)_{1<i\leq n}$ its dual family, then $f = \sum_i w_i \psi_i$ with $(w_i)_{1<i\leq n}$ the family of weight of connections (Zaied, Jemai and Ben Amar, 2008).

A weight can be calculated using the dual wavelet: $w_k = < f, \tilde{\psi}_k >$

$$
\begin{aligned}
< f, \tilde{\psi}_k >&= \int f(x)\tilde{\psi}_k dx \\
&= \int \left[\sum_i w_i \psi_i\right] \tilde{\psi}_k dx \\
&= \sum_i w_i \int \psi_i \tilde{\psi}_k dx \\
&= \sum_i w_i \delta_{i,k} \\
&= w_k
\end{aligned}
\tag{2}
$$

For a successful calculation, we are led, at each step of the optimization process, to know the family of dual wavelet of our wavelet network. The family of dual wavelet $\tilde{\psi}$ is calculated using the following formula:

$$
\tilde{\psi}_i = \sum_{j=1}^{n} (\Psi_{i,j})^{-1} \psi_j \quad \text{with } \Psi_{i,j} =< \psi_i, \psi_j > \tag{3}
$$

To demonstrate that $\tilde{\psi}$ is dual to ψ, we must check biorthogonality condition.

$$
< \psi_i, \tilde{\psi}_i >= \delta_{i,j} \tag{4}
$$

$$
\begin{aligned}
&< \psi_i, \sum_{j=1}^{n} (\Psi_{k,j})^{-1} \psi_j >= \int \psi_i(t) \left[\sum_{j=1}^{n} (\Psi_{k,j})^{-1} \psi_j dt \right] \\
&= \sum_{j=1}^{n} (\Psi_{k,j})^{-1} \left[\int \psi_i(t) \psi_j dt \right] \\
&\quad = \sum_{j=1}^{n} (\Psi_{k,j})^{-1} < \psi_i, \psi_j > \\
&\quad = \sum_{j=1}^{n} (\Psi_{k,j})^{-1} \Psi_{j,i} \\
&\quad = \delta_{i,k}
\end{aligned} \tag{5}
$$

4 CORPUS

We have tested our algorithm on two corpuses:
The first corpus was recorded by 11 speakers (5 women and 6 men) from works of (Boudraa and Boudraa, 1998). We segmented this corpus manually by PRAAT to Arabic words and we chosen 13 different words. Training corpus was about 90% and test corpus 10%. The second algorithm was recorded by 14 speakers. It was about Tunisian city name.

We have chosen Mel-Frequency Cepstral Coefficients MFCC and Perceptual Linear Predictive PLP coefficients to represent acoustic data.

5 RESULTS

The following figures illustrate the recognition rate given by the intelligent algorithm and the algorithm that offer randomly selection of wavelet.
We tested our approach on Arabic word using MFCC and PLP coefficients.

Figure 3 illustrate recognition rate of intelligent and normal algorithm on Arabic words using MFCC coefficients.

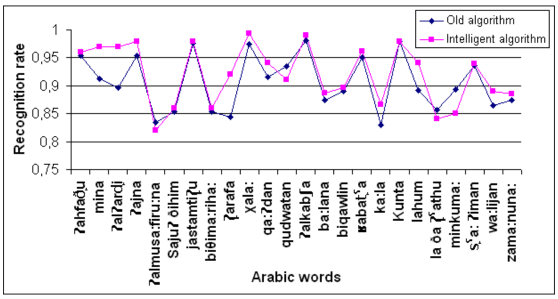

Figure 3: Evaluation of intelligent algorithm on Arabic words using MFCC coefficients.

Figure 4 illustrate recognition rate of intelligent and normal algorithm on Arabic words using PLP coefficients.

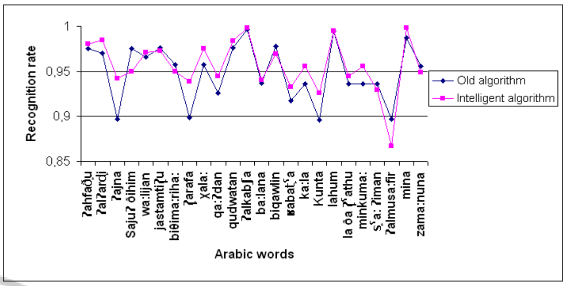

Figure 4: Evaluation of intelligent algorithm on Arabic words using PLP coefficients.

Figure 5 illustrate recognition rate of intelligent and normal algorithm on Tunisian city names using MFCC coefficients.

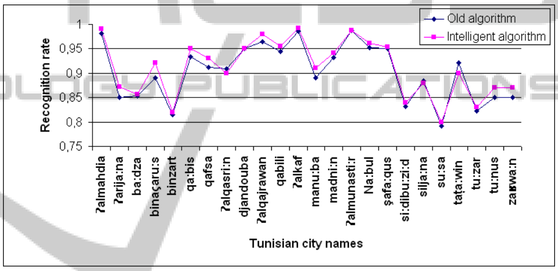

Figure 5: Evaluation of intelligent algorithm on Tunisian city names using MFCC coefficients.

Figure 6 illustrate recognition rate of intelligent and normal algorithm on Tunisian city names using PLP coefficients.

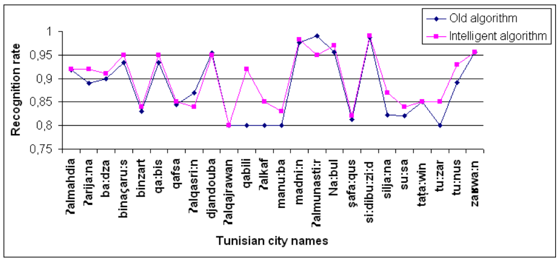

Figure 6: Evaluation of intelligent algorithm on Tunisian city names using PLP coefficients.

According to all the figures, we can say that the intelligent algorithm using intelligent selection of wavelet is better than the algorithm using randomly selection of wavelet

These results can be improved by increasing the size of the learning base, or by improving the quality of recordings…

6 CONCLUSIONS

Speech recognition, despite rising performance, was not able to reach expected results for large vocabulary applications, for real time applications and for real communication applications. Our paper contributes to the improvement of speech recognition systems by suggesting a new technique based on wavelet networks. A new type of modelling is setting forward with the birth of this new technique. Each acoustic unit is modelled by wavelet network refining the exposure of its characteristics.

Giving the finding of this new modelling technique, it could be adopted in large vocabulary applications and real applications aiming at an extreme performance.

REFERENCES

Ben Amar, C., Jemai, O., 2005. Wavelet Networks Approach for Image Compression, *International Journal on Graphics, Vision and Image Processing* GVIP, Vol. SI1, pp. 37- 45.

Szu, H., Telfer, B., Kadambe, S., 1992. Neural network adaptative wavelets for signal representation and classification. *Optical Engineering* 31:1907-1961.

Daugman, J., 2003. Demodulation by complex-valued wavelets for stochastic pattern recognition, *Int'l Journal of Wavelets, Multi-resolution and Information Processing,* vol. 1, no. 1, pp 1-17.

Zhang, Q., Benveniste, A., 1992. Wavelet Networks, IEEE *Trans. on Neural Networks* 889-898.

Bakis, R., 1976. Continuous speech recognition via centisecond acoustic states, In *Proc. 91st Meeting of the Acoustic Society in America*.

Ejbali, R., Benayed, Y., Alimi, A. M., 2009. Arabic continues speech recognition system using context-independent, 6th *International Multi-Conference on Systems, Signals and Devices*, Jerba Tunisie.

Yengar, S. S., Cho, E. C., Phoha, V., 2002. Foundations of Wavelet Networks and Applications, Chapman and Hall/CRC Press, June 2002.

Boudraa, M., Boudraa, B., 1998. Twenty list of ten arabic Sentences for Assessment, ACUSTICA acta acoustica. Vol. 86, no. 43.71, pp. 870-882.

Zaied, M., Jemai, O., Ben Amar, C., 2008. Training of the Beta wavelet networks by the frames theory: Application to face recognition, The international Workshops on Image Processing Theory, Tools and Applications, Tunisia.

Kruger, V., Sommer, G., 2001. Gabor Wavelet Networks for object representation, Technical Report CS-TR-4245, University of Maryland, CFAR.

Pati, Y. C., Krishnaprasad, P. S., 1993. Analysis and Synthesis of Feedforward Neural Networks Using Discrete Affine Wavelet Transformations. IEEE *Trans. On Neural Networks*, Vol. 4, No. 1, pp.73-85.