CLASSIFICATION OF FHR DURING THE INTRAPARTUM PERIOD USING WAVELET NEURAL NETWORKS TRAINED BY PSO

G. Georgoulas*, C. Stylios** and P.P. Groumpos*

* Laboratory for Automation & Robotics, University of Patras, Patras, Greece
** Dept. of Communications, Informatics and Management, TEI of Epirus, Kostakioi, Artas, Greece

georgoul@ee.upatras.gr

Abstract: In this paper we present a novel method for the discrimination of fetuses suspicious of developing acidemia from healthy ones. The proposed methodology employs wavelet analysis, neural networks and a newly developed paradigm from the field of evolutionary computation in a unified framework, to achieve better feature extraction and classification results for Fetal Heart Rate. The methodology is tested in experimental data set and the discrimination results are promising paving the way for further investigation and experimentation.

Introduction

Electronic Fetal Monitoring (EFM), usually named cardiotocography, has been widely used for antepartum and intrapartum fetal surveillance. EFM refers to the continuous recording and monitoring of Fetal Heart Rate (FHR) and Uterine Activity (UA), also known as cardiotocogram (CTG), which is depicted in Figure 1. In daily obstetric practice, obstetricians largely rely on information from the FHR. During the final period of labor and especially during the stressful delivery process, the risk of developing fetal hypoxia is increased. Monitoring of FHR is extensively used as an indirect screening test on fetal acid base balance [1].

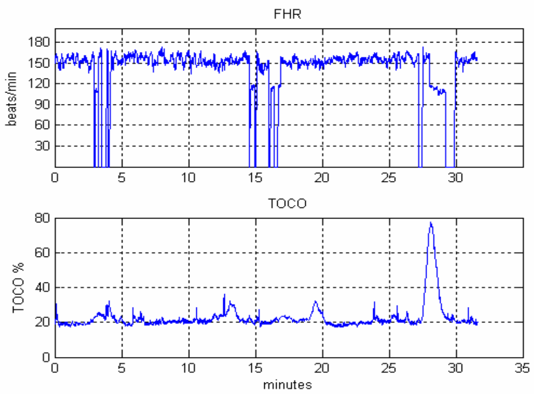

Figure 1. Typical CTG, with the FHR in the upper part and the UA in the lower part.

The inconsistency in interpretation [2] and the increase of false positive diagnosis [3] have raised the issue of whether a reliable and reproducible interpretation of the FHR patterns can be developed. The solution to this issue may be the deployment of new methodological tools [4-16], considering new indices more responsive to normal and pathological fetal conditions.

A quite novel signal processing method is the wavelet analysis. The way wavelet analysis localizes signal's information in the time-frequency (time-scale may be a more appropriate term), making it especially suitable for the analysis of non-stationary signals as an alternative to the classical short-time Fourier transform. Wavelet analysis has been already used as a tool for the extraction of scale-dependent and time scale dependent features for the classification of the FHR during the intrapartum period [17, 18].

However, those approaches use predefined features and they have not any adaptation ability. On the other hand, neural networks are learning paradigms that are naturally designed to adapt using a training set and without any other prior knowledge concerning the problem in hand. The idea of combining both wavelets and neural networks has resulted in the formulation of wavelet networks – a feedforward neural network with one hidden layer of nodes, whose basis functions are drawn from a family of orthonormal wavelets [19, 20]. Various algorithms can be used for training, such as conjugate gradient method, stochastic gradient algorithm, genetic algorithms etc. Most of applications of wavelet neural networks consider the regression problems [19, 21].

Particle swarm optimization (PSO) is a stochastic, population-based optimization algorithm [22]. It belongs to the class of swarm intelligence algorithms, which are inspired from the social dynamics and emergent behavior that arises in socially organized colonies. It is a derivative free optimization method [23] and has been used with reasonable success in many applications including the training of neural networks [22].

In this work we apply PSO for training a wavelet neural network that will be able to discriminate fetuses suspicious of developing acidemia from those that are coping well with the stress induced during labor.

ISSN: 1727-1983 © 2005 IFMBE

Materials and Methods

Wavelet neural Networks

Artificial neural networks (ANNs) are increasingly used in problem domains involving classification. They are adept at finding commonalities in a set of seemingly unrelated data and for this reason are used in a growing number of classification tasks.

The wavelet transform is a decomposition of the original signal onto a set of basis functions called wavelets. Those basis functions are obtained from a single prototype wavelet, which is referred to as the “mother wavelet” $\psi(t)$, by dilations and contractions (scalings), as well as shifts:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{1}$$

Wavelet network (WN) tries to combine aspects of the wavelet transformation for feature extraction and selection purposes with the characteristic decision capabilities of neural network approaches [24]. A WN can be described as an expanded perceptron with so-called wavelet nodes as preprocessing units for feature extraction (Figure 2). Each wavelet node computes the inner product of the predefined wavelet and the signal under investigation. The nodes are described by a shift parameter, b_k, and a scale parameter, a_k and the output of each node is

$$\phi_k = \frac{1}{\sqrt{a}}\int s(t)\psi^*\left(\frac{t-b_k}{a_k}\right) \tag{2}$$

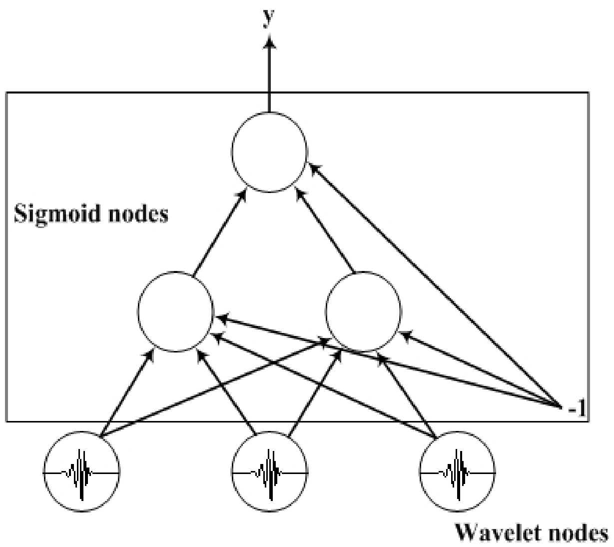

Figure 2. A simple wavelet neural network.

Figure2 shows a simple wavelet neural network with 3 wavelet nodes in the input layer, 2 sigmoid nodes in the hidden layer and one sigmoid node in the output layer. The part of the network contained in the rectangular is a typical perceptron with sigmoid activation functions.

The parameters related to the wavelet nodes, as well as the weights, are tuned using a supervised learning scheme. Learning is often formulated as the minimization of an error function, such as the total mean square error between the actual output and the desired output summed over all available data. While the sum-of-squares error function is appropriate for regression, for classification problems it is often advantageous to optimize the network using the cross entropy error function [25]

$$E = -\sum_{n=1}^{N}\left\{t_n \ln y_n + (1-t_n)\ln(1-y_n)\right\} \tag{3}$$

where N is the total number of training patterns, $t_n \in \{0,1\}$ is the label of the n^{th} pattern and y_n is the output of the neural network when presented with the n^{th} pattern.

Particle Swarm Optimization

PSO is a non-linear method which falls under the class of evolutionary computation techniques. It was originally proposed by J. Kennedy as a simulation of social behavior, and it was initially introduced as an optimization method in 1995 [26].

PSO is a population based method, i.e., it exploits a population of individuals to probe for promising regions of the search space, simultaneously. The population is called a swarm and the individuals (i.e., the search points) are called particles. The movement of the particles is stochastic; however it is influenced by the particle’s own memories as well as the memories of its peers.

A minimization (or maximization) of the problem topology is found both by a particle remembering its own past best position and the entire swarm’s best overall position. The Basic PSO algorithm consists of the velocity and position equation:

$$v_i(t+1) = \phi(t)v_i(t) + \eta_1 r\left[p_i - x_i(t)\right] + \eta_2 r\left[p_{g(t)} - x_i(t)\right] \tag{4}$$

$$x_i(t+1) = x_i(t) + v_i(t) \tag{5}$$

where i is the particle index, $v_i(t)$ is the current velocity of the i^{th} particle, $\phi(t)$ is an inertia function (usually a linearly decreasing one), $x_i(t)$ is the current position of the i^{th} particle, p_i is the position with the best fitness value visited by the i^{th} particle, $g(t)$ is the particle with the best fitness among all the particles (best position found so far - global version of the pso [22]), r is a positive constant called acceleration constant and η_1, η_2 are random numbers uniformly distributed in [0, 1].

ISSN: 1727-1983 © 2005 IFMBE

In our case the fitness function is the cross-error function and each dimension of a particle corresponds either to the connecting weights or the translation and dilation parameters of the wavelet nodes.

Experimental set up

The experimental data set consists of 40 FHR signals. The FHR signal are divided in two 2 subsets depending on whether the fetus has developed acidemia or not. Acidemia was determined for this work based on the pH value of umbilical artery (measured just after the delivery). The boundary to discriminate fetus is set to 7.1. Therefore, in the first subset, we included those FHR signals that belonged to fetuses with umbilical artery blood pH less than 7.1 and in the second subset, those FHRs signals that belonged to fetuses with umbilical artery blood pH more than 7.2. In the data set we didn't include fetuses with umbilical artery pH in the range [7.1, 7.2]. All FHR records had been acquired during the final stage of the labor and, in fact, as close as possible to delivery. This means that the data sets were time-biased free and a direct association could be made between the segment of the signal used and the fetal outcome. The recordings had durations ranging for 20 minutes to more than 1 hour. In this work we focused on FHR recordings as close as possible to delivery and for segments of relatively small duration

FHR is a very noisy signal with a lot of spiky artifacts and even periods of missing data due to the movement of the baby and the stress induced during the labor, leading to the displacement of the transducer used for its acquisition. This kind of noise cannot be eliminated in the source and it is always present in cardiotocographic records. Therefore, before any further processing, it is necessary to eliminate the noise from FHR, so we implemented a noise removal algorithm for FHR [11].

After the preprocessing stage, 5 minute segments were extracted from the end of each of the recordings. The wavelet neural network operated on these segments with its input wavelet nodes. In this work we tried to keep the structure of the wavelet network as simple as possible. Therefore we tried a network with 3 wavelet nodes, 2 sigmoid nodes in the hidden layer and one sigmoid node in the output layer. We also considered dyadic wavelet nodes, i.e. we allowed only discrete predefined steps for both the shift parameter, b_k , and the scale parameter, a_k leading to the following wavelet representation:

$$\psi_{m,n}(t) = \frac{1}{\sqrt{2^m}}\psi\left(\frac{t-n2^m}{2^m}\right) = 2^{-m/2}\psi\left(2^{-m}t-n\right) \qquad (6)$$

The parameters *n, m* were selected as part of the optimization process involving the PSO method. Each wavelet node takes as input the FHR signal and produces a wavelet coefficient which is then forward propagated through the part of the network that is in fact a perceptron with sigmoid activation functions.

Results

To evaluate the performance of the classifier we divided the 40 cases into 4 (non-overlapping) subsets, each one consisting of 5 sets from the "normal" and 5 from the "risk" group. The SVM classifier was trained on all subsets except for one, and the validation performance was assessed on the subset left out. We repeated this procedure 5 times, each time using a different subset for testing. Since the PSO is a stochastic process the above procedure was repeated 5 times and the performance was assessed taking the average over the 5 iterations.

The power of the PSO in optimization can lead to overfiting in the case of neural network learning. We must keep in mind that PSO is an optimization method but the minimization of an error criterion does not guarantee good generalization ability of the neural network. Therefore we had to modify the error function including a weight decay penalty term [25], and thus the new fitness function was given by the following expression.

$$E_r = E + \lambda \sum w_i^2 \qquad (7)$$

The best results were achieved using Daubechies wavelets [27] with 4 vanishing moments. The results for the WN with and without weight decay are summarized in Table 1.

Table 1: Classification performance of the WN

	Overall accuracy	**accuracy (normal)**	**accuracy (normal)**
Without weight decay	58.50	60	57
With weight decay	77.50	84	71

Discussion

In this work we focused on the difficult task of discriminating between fetuses that are suspicious of developing academia and those that are coping well with the stress induced during labor. Our approach deployed a very simple wavelet neural network with only 3 wavelet nodes. This means that only 3 features were employed, which is the smallest amount of features that we have used so far [17, 18]. However, the results are comparable to those reported in [17], where 6 features based on isolated coefficients had been employed. This indicates the potential capabilities of the proposed method.

The inclusion of the weight decay term in the error function results in a dramatic improvement in the generalization performance of the wavelet neural network and more attention should be given in finding an optimum value for the regularization parameter λ.

Moreover, it must be mentioned that the threshold for the pH value could be probably chosen lower for the "hypoxic" case. A more justified threshold would be the value of pH at 7, but this would compromise more the classification performance, since only 2 cases would fulfill that criterion, leaving 34 to the normal set. It is obvious that with this partition, overfitting would occur. It is worth mentioning that only very low pH values (6.8) are related to neonatal death or major neurological damage [28].

Conclusions

This research work presents a novel methodology for feature extraction and discrimination of fetuses at risk to suffer from acidocis. A more elaborate analysis of data and more profound experimental work on learning of networks is necessary to in order to further validate the feasibility of this approach. In general, no single algorithm is an overall winner for all kinds of networks. The best training algorithm is problem dependent.

In future work we will also employ wavelet networks equipped with wavelet nodes with a finer resolution of the parameters b_k, a_k. Furthermore different neural network topologies will be used (Radial Basis Functions networks, networks with mixed activation functions etc.) combined with the layer of the wavelet nodes, along with different error functions in order to fully exploit the capabilities of this hybrid approach.

To sum up, the proposed method is promising but it still has to be tested using a bigger data set before safer conclusions can be drawn.

Acknowledgments

Authors would like to thank Prof. João Bernardes, Dep. Gynecology & Obstetrics, Porto Faculty of Medicine, Porto, Portugal for providing the FHR data collected within the Research Project POSI/CPS/40153/2001, from Fundação para a Ciência e Tecnologia, Portugal.

Funding for this research was provided by the program EPEAEK II: Archimedes -Research Support in TEI of the Ministry of National Education and Religious Affairs of Greece.

References

[1] GEIJN, H. P. V. (1996): 'Developments in CTG analysis', *Bailliers Clin. Obstet. Gynaecol.,* **10**(2), pp. 185-209

[2] BERNARDES J., PEREIRA A. C., CAMPOS D. A., GEIJN, H. P. V., and LEITE L. P. (1997): 'Evaluation of interobserver agreement of cardiotocograms', *Int J. Gynecol Obst.,* **57**(1), pp. 33-37

[3] McDONALD D., GRANT A., SHERIDAN-PEREIRA M., BOYLAN P., and CHALMERS I. (1985): 'The Dublin randomized controlled trial of intrapartum fetal heart rate monitoring', *Am. J. Obstet Gynecol.,* **152**, pp. 524-39.

[4] ARDUINI D., RIZZO G., PIANA G., BONALUMI A., BRAMBILLA P., and ROMANINI C. (1993): 'Computerized Analysis of Fetal Heart Rate: I. Description of the System (2CTG)', *J. Maternal Fetal Invest.*, **3**, pp. 159-163

[5] MANTEL R., GEIJN, H. P. V., CARON F. J. M, SWARTJES J. M., WOERDEN E. E. V., and JONGSMA H. W. (1990): 'Computer analysis of antepartum fetal heart rate: 1. Baseline determination', *Int. J. Biomed. Comput.,* **25**(2), pp. 261-272

[6] MANTEL R., GEIJN, H. P. V., CARON F. J. M, SWARTJES J. M., WOERDEN E. E. V., and JONGSMA H. W. (1990): 'Computer analysis of antepartum fetal heart rate: 2. Detection of accelerations and decelerations', *Int. J. Biomed. Comput.*, **25**(2), pp. 273-286, 1990.

[7] TAYLOR G. M., MIRES G. J., ABEL E. W., TSANTIS S., FARRELL T., CHIEN P. F. W., and LIU Y. (2000): 'The development and validation of an algorithm for real time computerized fetal heart rate monitoring in labour', *Br. J. Obstet. Gynaecol.,* **107**, pp. 1130-1137

[8] JEZEWSKI J., and WROBEL J. (1993): 'Foetal monitoring with automated analysis of cardiotocogram: The KOMPOR system', Proc. 15th Ann. Card. IEEE/EMBS, San Diego, CA, 1993, pp. 638-639

[9] MAEDA K. (1990): 'Computerized analysis of cardiotocograms and fetal movements', *Bailliers Clin. Obstet. Gynaecol.,* **4**(4), pp. 1797-813

[10] BERDINAS B. G., BETANZOS A. A., and ROMERO O. F. (2002): 'Intelligent analysis and pattern recognition in cardiotocographic signals using a tightly coupled hybrid system', *Artificial Intelligence*, **136**, pp. 1-27

[11] BERNARDES J., MOURA C., SA J. P. M., and LEITA L. P. (1991): 'The Porto system for automated cardiotocographic signal analysis', *J. Perinat. Med*., **19**, pp. 61-65

[12] DAWES G. S., MOULDEN M., and REDMAN C. W. (1995): 'Computerized analysis of antepartum fetal heart rate', *Amer. J. Obstet. Gynecol.,* **173**(4), pp. 1353-1354

[13] SKINNER J. F., GARIBALDI J. M. and IFEACHOR E. C. (1999): 'A Fuzzy System for Fetal Heart Rate Assessment', Proc. of the 6th Fuzzy Days Conference, Dortmund, Germany, 1999, pp. 20-29

[14] MAGENES G., SIGNORINI M.G., and ARDUINI D. (2000): 'Classification of cardiotocographic records by neural networks', Proc. of the IEEE-INNS-ENNS International Joint Conference, **3**, pp. 637-641

[15] SALAMALEKIS E., THOMOPOULOS P., GIANARRIS D., SALLOUM I., VASIOS G., PRENTZA A., and KOUTSOURIS D. (2002): 'Computerised intrapartum diagnosis of fetal hypoxia based on fetal heart rate monitoring and fetal pulse oximetry recordings utilising wavelet analysis and neural networks', *Br. J. Obstet. Gynaecol.,* **109**(10), pp. 1137-1142

[16] CUNG T. K. H., MOHAJER M. P., YANG X. J., CHANG A. M. Z., and SAHOTA D. S. (1995): 'The prediction of fetal acidosis at birth by computerized analysis of intrapartum cardiotocography', *Br. J. Obstet. Gynaecol.,* **102**, pp. 454-460

[17] GEORGOULAS G., STYLIOS C. D., and GROUMPOS P. P. (2005): 'Investigation and comparison of different scale dependent features for fetal heart rate classification', Proc 16th IFAC World Congress, Prague, Czech Republic, 2005

[18] GEORGOULAS G., STYLIOS C. D., and GROUMPOS P. P. (2005): 'Classification of fetal heart rate using scale dependent features and support vector machines', Proc 16th IFAC World Congress, Prague, Czech Republic, 2005

[19] ZHANG Q., and BENVENISTE A. (1992): 'Wavelet Networks', *IEEE Trans. Neural Networks*, **3**(6), pp 889-898

[20] IYEBGAR S. S., CHO E. C., and PHOHA V. V. (2002): 'Foundations of wavelet networks and applications', (Chapman & Hall/CRC)

[21] ZHANG J., WALTER G.G., MIAO Y., and LEE W.N.W. (1995): 'Wavelet Neural Networks for Function Learning', *IEEE Trans. Signal Processing*, **43**, 6, pp 1485-1497

[22] KENNEDY J., and EBERHART R. C. (2001): 'Swarm Intelligence', (Morgan Kaufmann)

[23] PARSOPOULOS K. E., and VRAHATIS M. N. (2002): 'Recent Approaches to Global Optimization Problems through Particle Swarm Optimization', *Natural Computing*, **1**, pp 235-306.

[24] DICKHAUS H., and HEINRICH H. (1996): 'Classifying Biosignals with Wavelet Netwrks: A Method for Noninvasive Diagnosis', *IEEE Engineering in Medicine and Biology*, pp 103-111

[25] BISHOP C. M. (1995): 'Neural Networks for Pattern Recognition', (Oxford University Press, New York)

[26] EBERHART R. C. and KENNEDY J., (1995): 'A new optimizer using particle swarm theory', Proc. 6th Symposium on Micro Machine and Human Science, pp. 39-43

[27] DAUBECHIES I. (1994): 'Ten Lectures on wavelets'. (SIAM, Philadelphia)

[28] PARER J. T. (1997): 'Handbook of fetal heart rate monitoring', (W. B. Saunders Company)

ISSN: 1727-1983