Journal of ELECTRICAL ENGINEERING, VOL. 55, NO. 5-6, 2004, 156–160

CHANNEL EQUALISATION USING A SOFT BACK–PROPAGATION LEARNING ALGORITHM

José D. Martín-Guerrero — Luis Gómez-Chova — Gustavo Camps-Valls — Antonio Serrano-López — Joan Vila-Francés — Javier Calpe-Maravilla — Emilio Soria-Olivas *

This paper proposes some useful modifications to the Expanded Range Approximation (ERA) learning algorithm. In channel equalisation, it is a common practise to recover the original signal using an artificial neural network. The ERA algorithm is an alternative to the usual backpropagation algorithm that mitigates the effect of local minima during the training process. In the basis of ERA, we propose a soft-based profiling of the homotopy parameter in order to avoid local minima in the error surface more efficiently than in ERA. We use three specific membership functions that allow to control the smoothness of the learning process, yielding better results and lower computational cost.

K e y w o r d s: artificial neural networks, learning algorithms, soft computing, channel equalisation

1 INTRODUCTION

Channel equalisation is a major problem in digital communications, specially in burst time division multiple access (TDMA) transmissions, where transmitting an equalising sequence in each burst is required [1]. When a binary signal is transmitted through a real dispersive channel, the received signal is affected by inter-symbol interference (ISI). Moreover, if noise is present, further corruption ensues. Therefore, in many practical cases, equalisation is necessary to recover the information from the received signal. Under adverse conditions such as low signal to noise ratio or when the distribution of the received samples is not linearly separable, non-linear techniques are most suitable, and artificial neural networks (ANN) have become a common choice [2].

Artificial neural networks realised in the form of a multilayer perceptron (MLP) and trained by the backpropagation (BP) learning algorithm have found wide application in this field [3]. However, the presence of local minima on the error surface of the BP algorithm can cause lock-up to non optimal solutions. This issue has been extensively reported in the literature and becomes evident both in linearly separable [4,5] or non-linearly separable [6,7,8] mappings. In addition, second-order classical minimization algorithms, such as conjugate gradient [9] or the quasi-Newton method [10] present the same problem.

A promising method to mitigate the effect of local minima in gradient-descent algorithms is the Expanded Range Approximation (ERA) algorithm [11]. In this paper, we propose useful modifications to the ERA learning algorithm. Our proposal is tested in TDMA channel equalisation, where the network is trained to perform binary classification.

The paper is organized as follows. In section 2, we review the ERA algorithm and present the modifications proposed. Experimental results in channel equalisation problems are shown and discussed in section 3. Finally, we give some conclusions and further work in section 4.

2 DESCRIPTION OF THE LEARNING ALGORITHM

2.1 Expanded Range Approximation

Even though the multilayer perceptron (MLP) trained by the backpropagation (BP) algorithm has been widely used as a classification tool, it has some drawbacks. One of the most important problems is that many local minima are likely to appear in the error surface when the complexity of the problem is high. An interesting possibility to circumvent this problem is based on the Expanded Range Approximation (ERA) algorithm [11]. In this algorithm, the training set for the network is progressively expanded from an initial average setting to the original complete training set in each training epoch. The expansion process generates different sets of training data and is controlled by a homotopy parameter $\lambda \in [0,1]$. If, for simplicity, we assume a network structure with a single output neuron, the expansion of the training set as a function of λ, is given by:

$$\hat{d}_p(i) = < d > + \lambda(i)\,[d_p - < d >] \tag{1}$$

where d_p is the desired signal for the p^{th} pattern, $\hat{d}_p(i)$ is the one used during the training, i is the training epoch, and $< d >$ denotes the arithmetic mean of the original P training patterns.

Initially, $\lambda = 0$ and $\hat{d}_p(1) = < d >$, where $\hat{d}_p(1)$ is the initial setting of the training set for the desired signal of the p^{th} pattern. This training set is then expanded after

* Department of Electronic Engineering, Digital Signal Processing Group (GPDS, http://gpds.uv.es) University of Valencia, Spain, E-mail: luis.gomez-chova@uv.es

ISSN 1335-3632

each epoch, in a step-wise manner, until $\lambda = 1$, when the training set is identical to the original set of input–output pairs.

The ability of the ERA algorithm to avoid local minima of the error function is based on three assumptions [11]:

- The problem defined by $\lambda=0$ has only one global minimum.
- The first small step ($\lambda << 1$) away from the $\lambda=0$ solution keeps the system in a global minimum.
- The range can be progressively expanded to $\lambda=1$ without displacing the system from the global minimum at any step.

These three points are satisfied for the XOR problem [11], however ERA can break down in more complex situations. A reason for this is that no general strategy has been found for the variation of the step size during the course of the expansion of the training set so as to assure global convergence, and thus it is very easy to fall into a local minimum. This is a major problem which may depend on the training set, the network architecture, and the initialisation of the synaptic weights.

We propose useful modifications to the ERA learning algorithm in order to alleviate this problem. An appropriate profiling of the homotopy parameter λ can avoid local minima in the error surface more efficiently than ERA thus showing *a priori* better generalisation capabilities.

2.2 Soft expansion in ERA

The variation of the training set within the ERA algorithm can be considered from a soft perspective. While ERA works in a step-wise manner, i.e. each value of λ is fixed a certain number of epochs, our approach is based on a continuous change in the value of λ, thus enhancing the potential ability to avoid local minima. In this way, we propose the soft ERA (SERA) algorithm in which the λ evolution can be carried out by using three different functions: triangular, Gaussian and bell-shaped, which are shown in Fig. 1.

The modified algorithms are denoted as SERA1 (Triangular), SERA2 (Gaussian), and SERA3 (Bell-shaped) with corresponding values of λ given by:

$$\begin{aligned} \lambda_1(i) &= i/n \\ \lambda_2(i) &= e^{-\left(\frac{i-c}{a}\right)^2} \\ \lambda_3(i) &= \frac{1}{1+\left|\frac{i-c}{a}\right|^{2b}} \end{aligned} \tag{2}$$

where i denotes the training epoch and n the final number of epochs. The parameters a, b, and c characterise the Gaussian and bell-shaped functions.

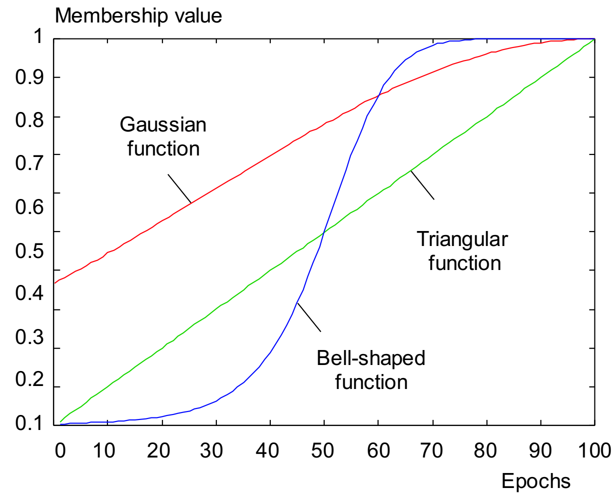

Fig. 1. Soft membership functions to describe λ parameter in the SERA algorithm as a function of the training epochs. Gaussian (dashed), triangular (solid) and bell-shaped (dots) functions are plotted.

By using these functions, the λ parameter is driven from zero, or from very low values, to 1 as the training progresses. In the case of the triangular function, the evolution of λ is linear with respect to the number of epochs. Since the variation in each epoch is very small, our proposal allows us to keep the model near the global minimum in a more easier way than that of a step-wise strategy.

The plotted Gaussian membership function yields a quasi-linear evolution of the parameter. In the first epochs, the evolution of λ is slower than that provided by the triangular function; in this way, it is easier to keep the system in a global minimum. After the initial epochs, the evolution is faster and thus, the training set is closer to the original one during more epochs.

The third alternative (the bell-shaped function) is similar to the Gaussian, but it provides the highest increase in the gradient of λ in the central epochs of the training process. As it begins with a value $\lambda = 0$, a global minimum at the beginning of the training process can be guaranteed and, since λ quickly approaches to 1, the training set is closer to the original one during more number of epochs in the training process than with the other membership functions. The latter feature allows us to get higher generalisation capabilities. However, the fast evolution in the central epochs of the training can take the system out of the global minimum.

Summarising, if the variation of λ is too smooth, local minima will be avoided, but the final network will be trained by using the real training set just in the last epochs, sacrificing its generalisation capabilities. In this way, the bell-shaped function is the most versatile option among the three proposed, since it has three parameters to select and thus allows us a better control on the gradient of λ during all the training process.

All these hypotheses about the function behaviour were tested by studying changes in the gradient of the error function. However, when dealing with complex problems, a small change of λ value can shift the problem from a global minimum to a local one.

3 METHODOLOGY AND RESULTS

3.1 Experimental setup

The system that we shall consider throughout this paper is depicted in Fig. 2. The input to the channel is assumed to be a sequence, $\{y_k\}$, of independent symbols extracted from a specific alphabet. The channel output is corrupted by random noise, $\{n_k\}$, which is considered to be an additive white Gaussian process. The equaliser task is to recover the transmitted input sequence, $\{y_k\}$, from the received sequence $\{r_k\}$.

Two well-known communication channels are used to assess the performance of the algorithms. The z-domain transfer functions of which are given by $H_1(z) = 0.5 + z^{-1}$ and $H_2(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$, which corresponds to real-like situations encountered in communications systems, as reported in [12].

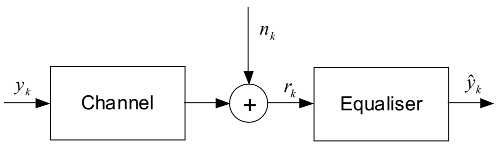

Fig. 2. Schematic of a data transmission system

At the receiver, a block of m received samples, $[r_k, r_{k-1}, , \dots , r_{k-m+1}]$, where k is the present discrete time index, is used to estimate the transmitted signal. A possible way to solve the problem is to consider that the set of received samples has to be classified according to the two possible classes (states) of the input: ± 1. Figure 3 shows an example of the receiver task when $m = 2$, plotting the output vectors for channels H_1(z) and H_2(z) when the transmission is corrupted with additive Gaussian noise with a standard deviation $\sigma = 0.15$. The x-axis shows the signal received in the instant n, and on the y-axis, the previous sample. It is clear that both classes, corresponding to ± 1, are not linearly separable.

3.2 Model comparison

We present results for the equalisation of the two previous communication channels, corrupted with additive Gaussian noise, by using different classification algorithms. The five algorithms are the classical MLP trained with the BP learning algorithm, the original ERA algorithm, and the three proposed SERA versions.

In order to equalise each channel, we used a training set formed by 250 patterns and a validation set constituted by 250 more. Finally, we evaluated the obtained Bit Error Rate (BER) for each Signal-To-Noise Ratio (SNR) value on 10^5 received symbols (test set), which represents a reasonable confidence margin for the least estimated error. All networks were initialised with the same synaptic weights. The learning rate and the number of neurons in the hidden layer were fixed after experimental studies to achieve the best results (five neurons in the hidden layer). We considered situations with and without additive noise ($\sigma_t = 0.15$ and $\sigma_t = 0$, respectively); both for training and validation sets. This methodology has been followed since in a real situation, the aim is to recover the original signal independently of the level of transmission noise.

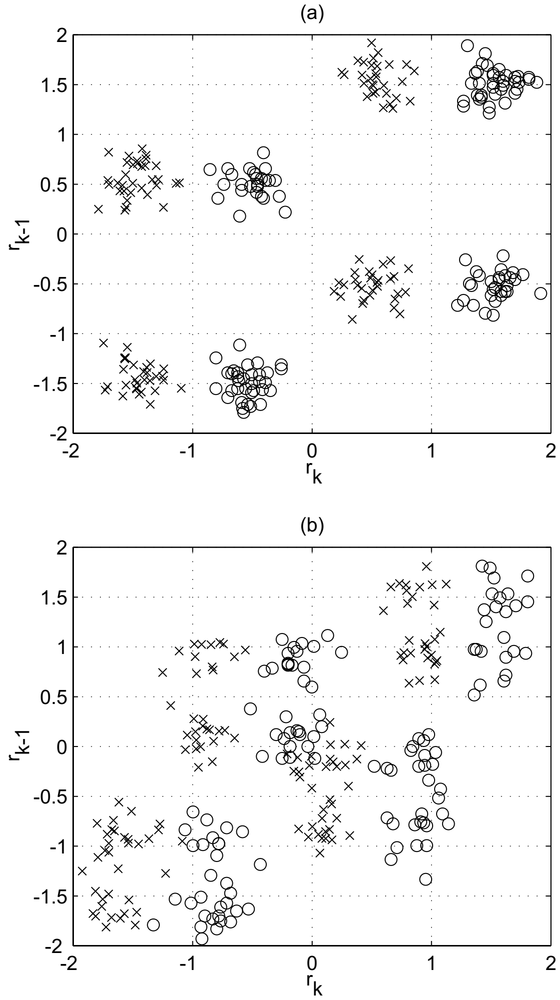

Fig. 3. Schematic of a data transmission system

Figure 4 depicts results in the test set over 10 realizations of the experiment. We use BER as a factor of merit, the lower the value of BER, the better is the equalisation carried out by the algorithm. Moreover, we study the dependence between BER and SNR since it is important to know the performance of the algorithms in different noisy environments. From empirical evidence, the parameters of λ functions shown in (2) are chosen as $c = n$, $a = n/2$, and $b = 4$. Figure 4 shows that the proposed SERA algorithms lead to better outcomes than those obtained with the original ERA algorithm, which are, in turn, much superior to those obtained with the classical BP algorithm.

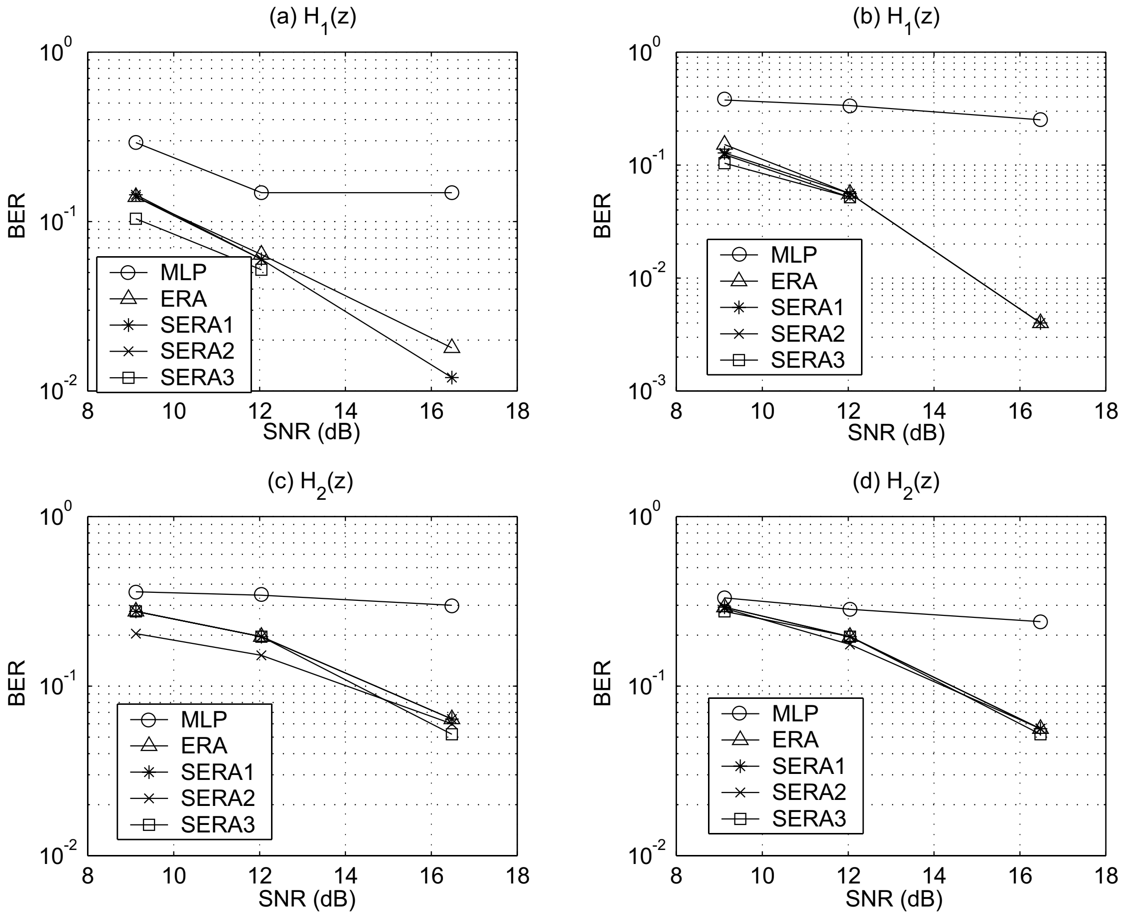

Fig. 4. Obtained Bit Error Rate (BER) for each Signal-To-Noise Ratio (SNR) for channel $H_1(z)$ with (a) $\sigma_t = 0$, (b) $\sigma_t = 0.15$, and channel $H_2(z)$ with (c) $\sigma_t = 0$ and (d) $\sigma_t = 0.15$.

This occurs for both channels and controlled noise training situations. Despite results with different SERA functions are quite similar, Bell-shaped membership function achieved slightly better results in most of the cases.

As it was expected, the value of BER is lower in the case of the first channel than in the second one, since it is easier for the algorithm to equalise its transfer function. It is also remarkable that a considerable dependence on the value of SNR appears for the case of ERA and SERA algorithms. Nevertheless, even in bad conditions of SNR, the BER is smaller than that achieved by the classical BP. The distance between BP and the other algorithms becomes much more significant when SNR is high. This suggests that even in excellent conditions of SNR the BP algorithm has its capability of equalisation limited. It is an obvious advantage of the ERA and SERA algorithms, which can perform much better in moderately noisy transmission channels.

It is remarkable that for the particular conditions of our problem, the ERA and SERA algorithms become cheaper in terms of computational burden. This is due to the fact that when dealing with desired signals whose value is $+1$ or -1, and the number of patterns belonging to each class is the same, equation (1) is trivially simplified.

4 CONCLUSIONS

In this paper, we have presented useful modifications to the ERA algorithm. The homotopy parameter of ERA is softly varied in a way inspired on the most common fuzzy-logic membership functions. More robustness to local minima and improved generalisation performance have been observed when Gaussian or Bell-shaped membership functions were employed. Simulation results have been presented for a channel equalisation application. Further work will consider adaptive update of membership function parameters.

Acknowledgement

This paper has been partially supported by the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2000-2003 (MCYT) under project DPI2001-2956-C02-01.

References

[1] PROAKIS, J. G. : Digital Communications, McGraw-Hill, New York, NY., 1995, 3rd edition.

[2] GIBSON, G. J.—SIU, S.—COWAN, C. F. N. : The Application of Nonlinear Structures to the Reconstruction of Binary Signals, IEEE Transaction on Signal Processing **39** No. 8 (1991), 1877–1884.

[3] HAYKIN, S. : Neural Networks: A Comprehesive Foundation, Prentice Hall, 1999.

[4] BRADY, M. L.—RAGHAVAN, R.—SHAWNY, J. : Back propagation fails when perceptron succeed, IEEE Transactions on Systems and Circuits **36** (1989), 665–674.

[5] SONTAG, E. D.—SUSSMAN, H. J. : Backpropagation can give rise to spurious local minima even for networks with no hidden layers, Complex Systems **3** (1989), 91–106.

[6] GORI, M.—TESI, A. : On the problem of local minima in back-propagation, IEEE Transaction on Pattern Analysis and Machine Intelligence **14** (1992), 78–86.

[7] SONTAG, E. D.—SUSSMAN, H. J. : Backpropagation separates where perceptrons do, Neural Networks **4** (1991), 243–249.

[8] McINERNEY, J. M.—HAINES, K. G.—BIAFORE, S.—HECHT-NIELSEN, R. : Error surfaces of multilayer networks can have local minima, UCSD. San Diego, CS89-157.

[9] KINSELLA, J. A. : Comparison and evaluation of the conjugate gradient method for efficient in feedforward neural networks with backward error propagation, Network 3 (1992), 27–35.

[10] BATTITI, R. : First and second methods for learning: between step-descent and Newton's method, Neural Computation **4** (1992), 141–166.

[11] GORSE, D.—SHEPERD, A. J.—TAYLOR, J. G. : The New ERA in Supervised Learning, Neural Networks **2** (1997), 343–352.

[12] KECHRIOTIS, G.—ZERVAS, E.—MANOLAKOS, E. S. : Using recurrent neural networks for adaptive communication channel equalization, IEEE Transactions on Neural Networks **5** (1994), 267–278.

Received 28 January 2004

José D. Martín-Guerrero (MSc) received his BSc in Physics and Electronics Engineering in 1997 and 1999 respectively, both from the Universitat de València. He is an assistant professor at the Department of Electronics Engineering.

Luis Gómez-Chova (MSc) received his BSc with first class honours in 2000 in Electronics from the Universitat de València, Spain. He is pursuing his PhD degree with a national research scholarship.

Gustavo Camps-Valls (PhD) received a BSc degree in Physics, a BSc degree in Electronics Engineering and a PhD degree in Physics from the Universitat de València. He is an assistant professor at the Department of Electronics Engineering.

Antonio Serrano-López (PhD) received a BSc degree in Physics and a PhD degree in Electronics Engineering from the Universitat de València. He is an assistant professor at the Department of Electronics Engineering.

Joan Vila-Francés (MSc) received his BSc with first class honours in Electronics from the Universitat de València, Spain. He is pursuing his PhD degree with a national research scholarship.

Javier Calpe-Maravilla (Prof, PhD) received his BSc with first class honours in 1989 and his PhD degree in 1993 in Physics from the Universitat de València. He is an Associate Professor and Head of the Department of Electronics Engineering.

Emilio Soria-Olivas (Prof, PhD) received his BSc in Physics with first class honours in 1992 and his PhD degree in 1997 in Electronics Engineering from the Universitat de València. He is an Associate Professor at the Department of Electronics Engineering.