ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 219-226

Discriminative learning for neural decision feedback equalizers

E.D. Di Claudio, R. Parisi and G. Orlandi

INFOCOM Dpt., University of Rome "La Sapienza"

Abstract

In this work new Decision-Feedback (DF) Neural Equalizers (DFNE) are introduced and compared with classical DF equalizers and Viterbi demodulators. It is shown that the choice of an innovative cost functional based on the Discriminative Learning (DL) technique, coupled with a fast training paradigm, can provide neural equalizers that outperform standard DF equalizers (DFEs) at practical signal to noise ratio (SNR). In particular, the novel Neural Sequence Detector (NSD) is introduced, which allows to extend the concepts of Viterbi-like sequence estimation to neural architectures. Resulting architectures are competitive with the Viterbi solution from cost-performance aspects, as demonstrated in experimental tests.

1. Introduction

Digital equalization of signals is required in any modern radio link or cellular telephony system. The decision-feedback equalizer is often employed, because it is simple and offers a better performance with respect to the linear equalizer [1]. A more complex approach to equalization is represented by the Viterbi algorithm, which exploits the knowledge of the mapping between transmitted and received sequences. However the channel model must be estimated from a preamble, which should be relatively short in order to maintain an adequate transmission efficiency. Estimation errors or channel miss-modelling (arising from non-linearities), Doppler effects and non-stationarities can impair the effectiveness of the Viterbi demodulator in real-world environments. Moreover, the Viterbi algorithm is much more expensive than the DFE, which precludes several high-speed and low-power applications. For these reasons, the DFE remains an attractive solution in many cases.

Neural nets provide powerful non-linear processing architectures, which have been proved to outperform traditional linear techniques in many common signal processing applications. In particular, neural networks have been proposed for digital equalization of communication channels [2][3][4].

Traditionally, the problem of equalization has been considered equivalent to the inversion of the transmission channel. Anyway, channel inversion is a problem in case of non-minimum phase channels, which are common in real world. An alternative and innovative approach considers equalization as a *classification* task [3][4],

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 219-226

consisting in establishing the correct mapping between received signals and target symbols. Linear and DF equalizers can be considered also as classifiers; however, being based on a (single-layer) linear combiner in the state space, they are limited in the shape and the complexity of the *decision regions* that they can form. Conversely, multilayer and recurrent neural networks are able to provide arbitrarily shaped decision contours [5] and are well known to yield very good supervised classifiers [8]. Anyway, drawbacks of neural networks are the high number of free parameters and the presence of non-linearities, that can slow down the optimization process. An important requirement is thus the use of fast learning strategies, able to guarantee safe global convergence and fast local descent to the optimum of the cost functional [6].

In the first part of this paper we describe a novel general approach to the equalization of digital signals, which satisfies the desired requirements of robust and fast classification. The proposed technique embodies the computational paradigm of neural nets in a DFE-like architecture. The new DF Neural Equalizer (DFNE) makes proper use of a powerful optimization technique (BRLS [6]), coupled with an error functional which has been proved to be particularly effective in classification tasks [7]. In the second part we introduce the Neural Sequence Detector (NSD), which incorporates the concepts of Viterbi-like sequence estimation into neural architectures for the equalization task. Performance comparisons with standard techniques for different channels demonstrate the validity of the proposed approach, expecially when the data model departs from assumptions and the computational cost is a critical issue.

2. Channel model

In fig. 1 a typical transmission system is depicted. $s(k)$ is the transmitted symbol stream, $n(k)$ is the additive noise, supposed zero-mean and Gaussian distributed. The channel may be non-linear, but the input-output symbol sequence map is assumed to be unambiguous. In modern interference-limited cellular telephony systems, the main error source is the Inter-Symbol Interference (ISI), rather than the thermal noise. The ISI consists in the spreading of symbol information through subsequent signal samples, and is the main problem in the relatively high SNR environment (10-20 dB), typical of most existing transmission systems.
The purpose of the equalizer is to estimate $s(k)$, minimizing the combined effects of ISI and noise. In particular, the DFE makes use of a set of delayed input samples and past detected symbols [1].

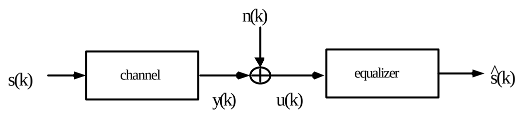

Fig. 1: typical transmission system.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 219-226

3. Neural Decision Feedback Equalizer

Recurrent Neural Networks (RNNs) can be successfully applied to the adaptive equalization of digital communication channels [4]. RNNs are able to yield significant performance when little information is available on the channel model. This fact can be explained by the very general assumptions made on the mapping from the received signal to the output symbol space, that recast the demodulation problem as a classification task.

The proposed neural network (depicted in fig. 2) is an evolution of the classical DFE and is considerably simpler and faster than existing structures, being composed of a two-layer perceptron. This architecture can be viewed as a RNN with an external feedback. Samples contained in both the input and the feedback tapped delay lines (TDLs) constitute the inputs to the first neuron layer. During the learning phase, the feedback TDL is fed by an internal replica of the transmitted preamble sequence. Then the switch commutes from position 1 to position 2 and the equalizer enters into the decision directed mode (DDE).

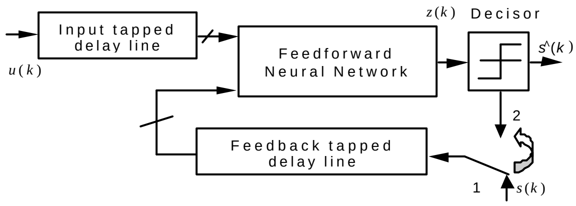

Fig. 2: Neural Decision Feedback Equalizer

The weight updating is made by the Block Recursive Least Squares (BRLS) algorithm described in [6]. This approach searches consistently for a local minimum of the error functional in a Newton-like fashion, thus allowing for a superlinear convergence rate. The choice of the cost functional should be related to the concept of equalization as a classification problem, where the objective is the separation of clusters generated by mapping the transmitted symbols through the channel input-output relationship. The usual choice of the Mean Squared Error (MSE) criterion might not fully satisfy the requirement for optimal classification.

Discriminative Learning (DL) was introduced in [7] in alternative to the MSE criterion as an enhanced tool for optimizing the boundary decision of non-linear classifiers. In the following we briefly recall the main concepts of DL.

Suppose that the aim of training is to associate an input pattern $\mathbf{u}(k)$ to one of M possible classes. As a first step, M discriminant functions $g_i(\mathbf{u}(k),\mathbf{w})$, depending on the network parameter vector $\mathbf{w}$, are introduced. The second step is the choice of an appropriate miss-classification measure, which is continuous with respect to the weights $\mathbf{w}$; a possible definition is the following:

$$d_i(\mathbf{u}(k)) = -g_i(\mathbf{u}(k),\mathbf{w}) + \left\{ \frac{1}{M-1} \sum_{i \neq j} g_i(\mathbf{u}(k),\mathbf{w})^p \right\}^{1/p} \tag{1}$$

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 219-226

This equation gives a measure of the classification error when the input belongs to the i-th class; in the simple case of two classes (BPSK), it reduces to the difference between the outputs, so $d_i(\mathbf{u}(k)) < 0$ means miss-classification, while $d_i(\mathbf{u}(k)) > 0$ implies a correct decision. The positive integer p is usually selected between one and two, depending on statistical properties of noise.
As a third step, the following error functional is defined as a function of the miss-classification measure:

$$l_i(\mathbf{u}(k),\mathbf{w}) = l_i(g(\mathbf{u}(k),\mathbf{w})) \qquad (2)$$

being l_i a differentiable zero-one "squashing" function, like a sigmoid or an exponential. The objective of learning is thus the minimization of the error functional l_i with respect to the weight vector $\mathbf{w}$, which can be performed by applying methods well-known in optimization theory.
This formulation allows to express the minimum classication error (or Bayes minimum risk) directly in terms of the functionals l_i when the discriminant functions $g_i(\mathbf{u}(k),\mathbf{w})$ give exactly the a posteriori probability of the i-th class given $\mathbf{u}$ [3]. This means that the minimum classification probability objective is conditioned on the choice of the correct discriminant functions. Anyway, due to their function approximation capabilities, NNs with the proper number of units are potentially able to converge to the true minimum Bayes risk [3]. In the present setup, the Bit Error Rate (BER), which is the commonly adopted performance index in a telecommunication system, is a measure of the miss-classificaton probability.
The DL has been already adopted in a practical recurrent neural equalizer [4]. The present architecture is much simpler than the fully recurrent architecture presented in [4], shares the same (or better) high adaptation speed (50-200 samples), but has better performance in terms of BER, expecially in the high-SNR region (10 dB or more) and with non-linear channels.

3.1 Simulation results

The proposed architecture was compared to a traditional DFE, enhanced by the use of the classical RLS adaptation algorithm in order to provide a fair comparison basis with the BRLS in terms of convergence rates and steady-state parameter misadjustment. The performance benchmark for BER was represented by a Viterbi decoder of depth five [1], which assumed the *exact* a priori knowledge of the channel model. The signal modulation was BPSK, and real arithmetic was employed throughout simulations.
It is worth to point out that practical performance of Viterbi demodulators can be significantly worse than in simulated environments, since the channel estimate is usually provided through only the few symbols of the preamble.
Several channels and architecture sizes were considered. Here we describe results obtained in two typical non-minimum phase and non-linear channels. For each channel, the length m of the input TDL, the length n of the feedback TDL and the

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 219-226

decision delay *d* [1] were selected according to the rules described in [3], and used for both the neural (NDFE) and the conventional DFE.
The NDFE had three input neurons and two output units, all featuring an hyperbolic tangent activaction function. Training was performed on 20,000 samples for all architectures, in order to ensure a steady-state performance index. After the training phase, the BER was computed by presenting new sequences of samples and averaging the results.
a) Linear non-minimum phase channel. Non-minimum phase channels are known to be difficult to equalize [1]. In this experiment we considered the non-minimum phase channel having the following transfer function in the *z* domain:

$$H(z^{-1}) = 0.3482 + 0.8704 \cdot z^{-1} + 0.3482 \cdot z^{-2} \tag{3}$$

The equalizer parameters were d=3, m=4, and n=2. Results in terms of BER versus SNR are summarized in fig. 3. The NDFE with DL outperforms the DFE in the high SNR region (SNR>14dB), which is the typical operating range of digital transmission systems.
b) Non-linear channel. Non-linear channels are met frequently in practice; for example, non-linearities can arise due to saturation phenomena of the amplifiers in the transmitter. Since usually Viterbi decoders assume a linear channel, a mismatching occurs on the symbol mapping in the presence of non-linearities. Furthermore, the identification of a non-liner channel is not a trivial task with the few samples typically available for training and synchronization. The resulting channel misadjustment can lead to a higher BER in both linear and non-linear environments. The proposed NDFE has, instead, a natural capability to cope with non-linearities.
In this second experiment we considered a simple non-linear channel, whose time difference equations are:

$$\begin{aligned} v(k) &= 0.3482 \cdot s(k) + 0.8704 \cdot s(k-1) + 0.3482 \cdot s(k-2) \\ y(k) &= v(k) \cdot [1 + 0.2 \cdot v(k)] \end{aligned} \tag{4}$$

All the equalizer parameters were the same as in the previous experiment. Results are shown in fig. 4. In this case the performance improvement begins at SNR=8 dB. Moreover, the Viterbi solution using the MSE gives a worse performance with respect to both DFEs.

4. Neural Sequence Detector (NSD)

The basic structure of the Neural Sequence Detector is the same of fig. 1, where the decision is made on the output pattern by a *soft Viterbi decoding*. At each time step, the algorithm minimizes the *Mean Square Error* (MSE) between the target (preamble sequence) and the detected outputs, and maximizes at the same time the Euclidean distance from the output sequence to all possible wrong codewords. The cost functional can be expressed as:

$$E(t_i) = -MSE_i + \frac{1}{N-1}\sum_{j=1, j\neq i}^{N} MSE_j \tag{5}$$

where MSE_i and MSE_j are respectively the output MSE with respect to the correct target sequence i and the generic (wrong) sequence j. This choice of the cost functional extends the concept of *Discriminative Learning* (DL) to a multidimensional output pattern [7]. The BRLS algorithm *maximizes* the cost (5) with respect to the network weights, giving superlinear and robust convergence [6]. In order to keep the functional $E(t_i)$ bounded, the non-linear activation functions placed at the outputs of the neural network must also be bounded (e.g. sigmoids) [5][8].

4.1 Decoding phase

In the decision phase a Viterbi-like decision criterion is implemented in the following way. At each time step, the network output pattern should be equal to the true (transmitted) sequence. On the basis of this argument, the actual network output is compared with a target which is dinamically formed and stored in the *feedback register*: the oldest target symbol exits the register and constitutes the actual detected symbol, while the remaining components are time-shifted. A new symbol is appended to the target; for all possible symbol choices, the output MSE is computed and ranked in a nonincreasing manner. The P targets with the lowest MSE are the *survivors* that are retained for the subsequent step and generate a *trellis* [1]. The best matching target is used for the optional weight tracking, performed by eqn. (5). In the steady state, with P symbols, L outputs and S survivors, $P \times S$ L-dimensional targets are formed at each time instant.

4.2 Experimental results

The proposed equalizer structure has been implemented using either a feedforward or a fully recurrent (Elman-type architecture [8]) neural network, and compared to the Viterbi algorithm (which constitutes the benchmark) in different environments. It is understood that neural architectures might exhibit some advantages when the noise is not Gaussian, the channel is non-linear, and co-channel interference is present. Each neural network had five input, eight hidden (or recurrent) and five output neurons. These architectures were found to be optimal for the channels considered. The modulation was BPSK [1].
The equation of the first test channel was:

$$\begin{aligned} u(n) &= -0.2052\cdot x(n) - 0.5131\cdot x(n-1) + 0.7183\cdot x(n-2) + \\ &+ 0.3695\cdot x(n-3) + 0.2052\cdot x(n-4) + \eta(n) \end{aligned} \tag{6}$$

where $x(n)$ is the transmitted symbol sequence, $u(n)$ is the received sequence and $h(n)$ is the white additive noise. This is a typical non-minimum phase channel. Fig. 5 summarizes the BER obtained for this channel in the presence of Gaussian noise. Fig. 6 shows the results when the noise belongs to a mixture of a Gaussian (60%) and a Laplacian (40%) distribution.
The second channel was non-linear; its equations were:

$$\begin{aligned} w(n) &= -0.2052 \cdot x(n) - 0.5131 \cdot x(n-1) + 0.7183 \cdot x(n-2) + \\ &\quad +0.3695 \cdot x(n-3) + 0.2052 \cdot x(n-4) \\ u(n) &= w(n) \cdot \left[1 + 0.2 w(n)\right] + \eta(n) \end{aligned} \tag{7}$$

Fig. 7 shows the results obtained in this case. In Figs. 5-7, cross-marks (x) are referred to the feedforward network, circles (o) indicate the recurrent network and stars (*) are used for the Viterbi equalizer.
It is important to remark the Viterbi decoder assumes the *perfect knowledge* of the channel, *even in the non-linear case*, while neural equalizers must estimate the channel. Anyway, the performance of the Viterbi algorithm are only slightly better, in the SNR range typical of cellular telephony.

References

[1] J.G.Proakis, *Digital communications*, McGraw-Hill, 1983.

[2] S. Sin, G.J. Gibson, C.F.N. Cowan, “Decision Feedback Equalisation Using Neural Network Structures and Performance Comparison with Standard Architecture”, *IEEE Proceedings*, Vol. 137, Part 1, No. 4, August 1990.

[3] S. Chen, B. Mulgrew, and S. Mc Laughlin, “Adaptive Bayesian Equalizer with Decision Feedback”, *IEEE Transactions on Signal Processing*, vol. 41, No. 9, September 1993.

[4] R.Parisi, E.D.Di Claudio, G.Orlandi and B.D.Rao, “Fast adaptive digital equalization by recurrent neural networks,” *IEEE Transactions on Signal Processing*, Special number on Neural Network applications, Vol. 45, No. 11, pp. 2731-2739, November 1997.

[5] G.Cybenko, “Approximations by superpositions of a sigmoidal function,” *Math. Control Signals Syst*., Vol.2, No.4, 1989.

[6] R.Parisi, E.D. Di Claudio, G. Orlandi and B.D. Rao, "A generalized learning paradigm exploiting the structure of feedforward neural networks," *IEEE Trans. on Neural Networks*, vol.7, no.6, November 1996.

[7] B.H.Juang, and S.Katagiri, “Discriminative learning for minimum error classification,” *IEEE Trans. on Signal Processing*, Vol.40, No.12, December 1992.

[8] S. Haykin, *Neural Networks: a comprehensive foundation*, Macmillan, New York, 1994.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 219-226

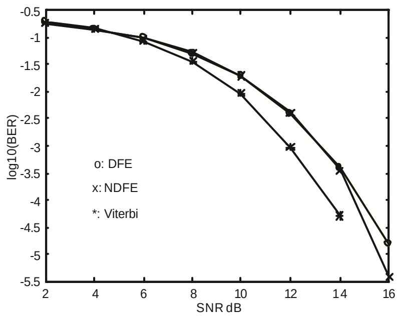

Fig. 3: Non-minimum phase channel: BER vs. SNR for DFE, NDFE and Viterbi

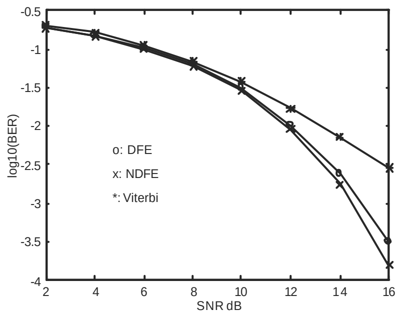

Fig. 4: Non-linear channel: BER vs. SNR for DFE, NDFE and Viterbi

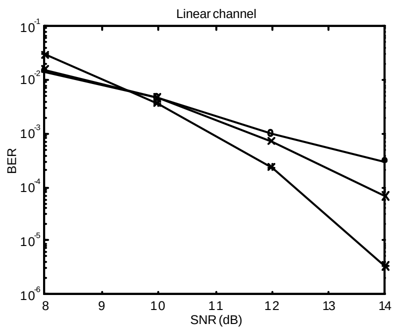

Fig. 5: NSDs and Viterbi equalizer, non-minimum phase channel and Gaussian noise

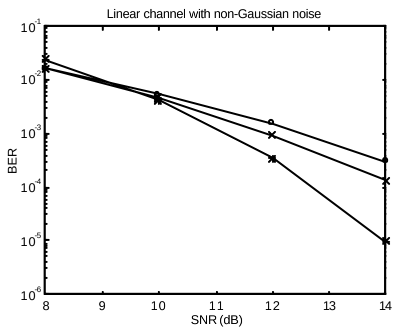

Fig. 6: NSDs and Viterbi equalizer, non-minimum phase channel and non-Gaussian noise

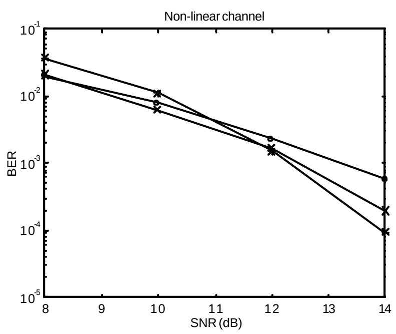

Fig. 7: NSDs and Viterbi equalizer, non-linear channel