NON-LINEAR EQUALIZERS THAT ESTIMATE ERROR RATES DURING RECEPTION

*Jesus Cid-Sueiro, Anibal R. Figueiras-Vidal**
Dpto. TSC e IT. Universidad de Valladolid
ETSI-Telecom. Campus Miguel Delibes, s/n,
47011 Valladolid, SPAIN
e-mail: jesus@tel.uva.es
Dpto de Tecnologías de las Com., Universidad Carlos III
Butarque, 15, 28911 Leganés, Madrid, SPAIN
e-mail: arfv@ing.uc3m.es

ABSTRACT

Neural Networks can be used to estimate the a posteriori probabilities of the transmitted symbols in digital communication systems. In this paper we apply this property to the on-line estimation of the bit error rate (BER) in the receiver, without using any reference signal. We discuss two different approaches to BER estimation: (1) computing the a posteriori symbol probabilities from estimates of the conditional distributions of the received data, and (2) estimating probabilities by gradient minimization of a special type of cost functions. We show that Importance Sampling (IS) techniques can be combined with the first approach to reduce drastically the variance of the probability estimates. Finally, we analyze the effect of channel variations during transmission.

1 Introduction

The application of neural networks to digital equalization has been a topic of discussion in the technical literature during the last years; many experiments have shown that neural equalizers obtain lower bit error rates (BER) than conventional structures in many cases; however, it seems not easy to do it keeping, at the same time, a moderate computational cost.

The Radial Basis Function (RBF) networks are one of the most efficient neural-network-based equalizers [1, 4]. They can compute optimal Bayesian decisions in channels with a finite distortion function; they can manage non-linear channels, they can be adapted to a non-Gaussian noise and, in a recursive form like that of Fig. 1, they can take optimal decisions based on all the received samples. However, they are not used in practical systems because other detectors, like those based on the Viterbi algorithm, have a similar performance and complexity [6]: in fact, recurrent RBF detectors [4] and sequence detectors are equivalent as the Signal to Noise Ratio (SNR) increases.

At the judge of the author, the main difference between both detection schemes is given by the kind of computations carried out before making the final decisions: the RBF-based equalizers can be used to compute the posteriori probabilities of the transmitted symbols; this information is not provided by the sequence detector. The advantages of using the posteriori symbol probabilities for blind equalization and tracking in time-variant environments have been discussed in several works [3, 4, 7].

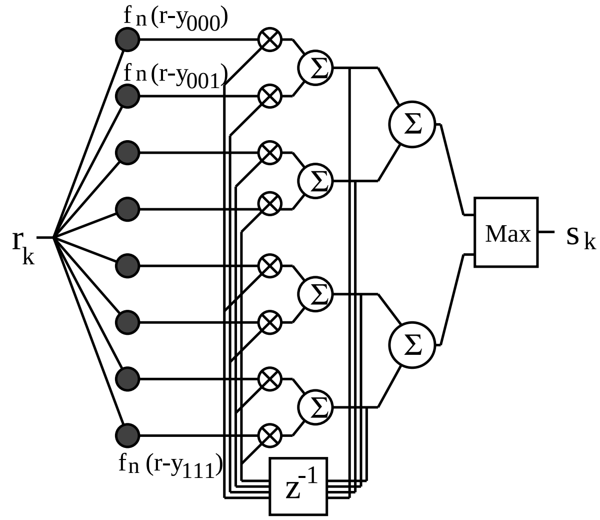

Figure 1: Structure of an optimal Recurrent RBF network (RRBF) for a finite response channel, r_k is the received sample and f_n is the noise probability density funtion. The number of nodes grows exponentially with the length of the channel response.

The ability to estimate posteriori probabilities is not exclusive of RBF-based detectors. It is well-known that, if a binary classifier is adapted to minimize the mean square error between its output and the class of the input, the network output is an estimate of the a posteriori probability of one of the classes. In [9, 5], general formulas for the cost functions leading to probability estimates are provided for binary and multi-class problems.

In this paper we show another advantage of computing symbol probabilities: the estimation of the BER in reception *without* any reference signal. BER estimates can be used by the receiver as a measure of the reliability of the data transmission process or even to control the transmission rate in variable rate digital modems.

2 BER estimation with RBF equalizers

Let us assume that a transmitter sends a sequence $\{s_k\}$ of N independent symbols which belong to alphabet $A = \{a_0, \ldots, a_{M-1}\}$. If the channel has finite memory m and the noise is additive, the received samples $\{x_k\}$ can be expressed as

$$x_k = h(s_k, \ldots, s_{k-m}) + n_k \tag{1}$$

where n_k are the noise samples and h is the channel distortion function, which may be non-linear. The optimal symbol-by-symbol detectors select the symbol which maximizes the "a posteriori" probability of being equal to the transmitted symbol: decisions $\{\widehat{s}_k\}$ are given by

$$\widehat{s}_{k-d} = \arg\left\{\max_i \{\Pr\{s_{k-d} = a_i \mid \mathbf{x}(k)\}\}\right\} \tag{2}$$

where d is the decision delay and $\mathbf{x}(k)$ is a vector containing the received samples used to make the current decision.

The RBF network can be trained in order to estimate the conditional probability density functions of the arriving samples, $f(\mathbf{x} \mid s_{k-d} = a_i)$. The posteriori probabilities of the transmitted symbols $P(s_{k-d} = a_i \mid \mathbf{x})$ can be computed from this applying elementary statistics. Since

$$P_{k-d} = 1 - \Pr\{s_{k-d} = a_i \mid \mathbf{x}(k)\} \tag{3}$$

is the error probability associated to decision $s_{k-d} = a_i$, it is immediate to show that P_{k-d} is an unbiased estimate of the overall symbol error probability of the detector, P_e:

$$\mathrm{E}\{P_{k-d}\} = P_e \tag{4}$$

In the following, P_{k-d} will be called an RBF estimator of the BER. An upper bound for the variance of the estimate can be easily derived

$$\sigma^2 = E\{P_{k-d}^2\} - E\{P_{k-d}\}^2 \tag{5}$$

Since $P_{k-d} < \frac{M-1}{M}$, $E\{P_{k-d}^2\} < \frac{M-1}{M} E\{P_{k-d}\} = \frac{M-1}{M} P_e$; therefore

$$\sigma^2 \leq \frac{\frac{M-1}{M} - P_e}{1 - P_e} \sigma_{mc}^2 \leq \sigma_{mc}^2 \tag{6}$$

where $\sigma_{mc}^2 = P_e(1 - P_e)$ is the variance of the Montecarlo (MC) estimate. Therefore, the variance of the proposed estimate is less than that of the MC method. This is, in general, a pessimistic bound. As an example, Fig. 2 illustrate the performance of the RBF estimate for the equalization of the linear channel $H_1(z) = 0.319 + 0.620z^{-1} + 0.634z^{-2} + 0.323z^{-3} + 0.087z^{-4}$, for different values of the Signal to Noise Ratio; the vertical axis represents the ratio between σ^2 and σ_{mc}^2 as a function

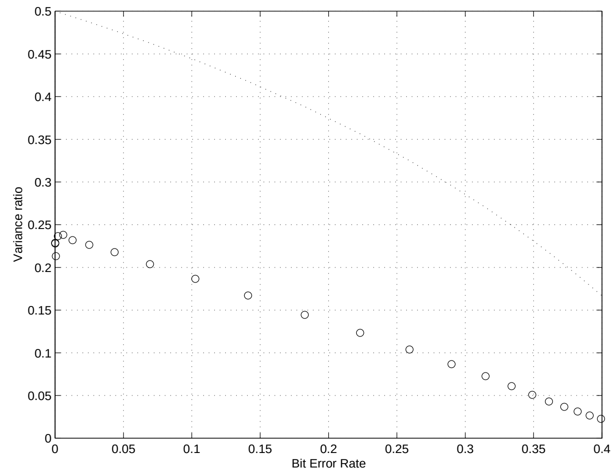

Figure 2: Variance ratio σ^2/σ_{mc}^2 vs SNR for the linear channel $H_1(z)$.The dotted line is the theoretical bound derived in the text

of the P_e. The continuous line represents the bound derived here.

Although other well-known BER estimation methods provide lower variance estimates [8], the RBF estimates have two major advantages: first, it does not require any knowledge about the transmitted symbol, second, it can be combined with Importance Sampling (IS) techniques [2]. Note that

$$\begin{aligned} P_e &= E\{P_{k-d}\} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P_{k-d}(\mathbf{x}) f_x(\mathbf{x}) d\mathbf{x} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P_{k-d}(\mathbf{x}) w(\mathbf{x}) f_x^*(\mathbf{x}) d\mathbf{x} \end{aligned} \tag{7}$$

where

$$w(\mathbf{x}) = \frac{f_r(\mathbf{x})}{f_r^*(\mathbf{x})} \tag{8}$$

Thus, if $\mathbf{x}$ follows the density $f_x^*(\mathbf{x})$, product $P_{k-d}(\mathbf{x})w(\mathbf{x})$ is an unbiased estimate of P_e. Thus, adding artificial noise to the received samples, we can modify the density function $f_x(\mathbf{x})$, reducing the number of samples that an accurate BER estimation requires. As an example, the variance of the RBF estimator for the linear channel $H_2(z) = 0.5 + z^{-1}$ with $SNR = 15dB$ can be reduced by a factor of 10^4 by artificially multiplying the noise variance by a factor of 4. The main disadvantage of the IS method is that it requires a second RBF structure to compute f_r^*. Also, as the posteriori probabilities are quotients of densities, IS methods are sensitive to inaccurate density estimates

3 BER estimates in time-variant channels

In practical situations, the channel response is not previously known by the user and must be estimated by the receiver. Moreover, the channel response may change during transmission, and the estimation algorithm must work during data reception, without any training sequence. As a result, the channel response is partially unknown and the channel estimate is not constant during time.

As discussed in [7, 4], the effect of the channel variations can be considered by modifying the noise variance estimate. From Eq.(1) we can write,

$$e_k = x_k - \widehat{h}(s_k, \ldots, s_{k-m}) = \Delta_h(s_k, \ldots, s_{k-m}) + n_k \quad (9)$$

where Δ_h is an error caused by an inaccurate estimate of h. RBF equalizers use the noise variance to compute conditional data densities. By computing them as the average of e_k^2, the channel variations can be taken into account. Fig.3 is an example of this; initially, a perfect knowledge of the linear channel response $H_1(z)$ (defined before), with SNR=18dB, was assumed. The channel response was modified at each symbol time according a simple random walk model

$$\mathbf{h}_k = \mathbf{h}_{k-1} + \mathbf{dh}_k \quad (10)$$

where $\mathbf{h}_k$ is a vector with the channel coefficients and $\mathbf{dh}_k$ is a random zero-mean Gaussian vector with variance σ_h^2. Decision-directed LMS was used for adaptive channel estimation,

$$\widehat{\mathbf{h}}_k = \widehat{\mathbf{h}}_{k-1} + \mu \widehat{e}_k \widehat{\mathbf{s}}_k \quad (11)$$

where $\widehat{\mathbf{s}}_k$ is a decision vector, and $\widehat{e}_k = x_k - \widehat{h}(\widehat{\mathbf{s}}_k)$.

The figure compares the RBF estimate with the real BER, for different values of the channel variance, σ_h^2. Note that, in some cases, the equalizer is not able to track the channel response and it fails in detecting the transmitted sequence; this is a known problem of decision-directed methods; other techniques has been proposed to reduce this problems (see [7], for instance); anyway, the adaptive algorithm is not the topic of this paper. The important thing is that, even in the cases where the channel is not accurately estimated, the equalizer obtains accurate estimates of the BER. This is essential to detect an undesired behaviour of the equalizer and re-start the training algorithm when it fails.

4 BER estimation with other neural networks

The RBF equalizer computes probabilities by computing, first, the conditional probability density function (pdf) of the arriving samples. The main advantage of the RBF equalizer is that it can compute the exact pdf of the samples, but it requires a finite channel response and Gaussian additive noise.

Alternatively, we can estimate the posteriori probabilities of the transmitted symbols without computing density functions. It is well-known that neural networks can estimate posteriori probabilities if the adequate cost function is minimized during learning: the mean square error or the cross entropy are some examples. In [9, 5], it is shown that any classifier adapted in order to minimize a cost function given by

$$C(\mathbf{y}, \mathbf{s}) = \sum_{i=0}^{M-1} \int_{s_i}^{y_i} g(\mu)(\mu - s_i) d\mu \quad (12)$$

where g an arbitrary non-negative function, is minimum when $y_i = P(s_i = 1 \mid \mathbf{x})$. The class of the input is specified by vector $\mathbf{s}$, which has all components equal to zero but a "1"; its position indicates the class. The square error and the cross entropy are included in the subset of cost functions given by

$$g(\mu) = \mu^\alpha \quad (13)$$

for $\alpha = 0$ and $\alpha = -1$, respectively. Therefore, any equalizer adjusted by minimizing these cost functions can be used to compute BER estimates..

We evaluated the BER estimates based on a "Softmax" perceptron, whose outputs are given by,

$$y_i = \sum_{j=1}^{n_f} y_{i,j} \quad (14)$$

where

$$y_{i,j} = \frac{\exp(\mathbf{w_{i,j}}^T \mathbf{x})}{\sum_{k=1}^{c} \sum_{l=1}^{n_f} \exp(\mathbf{w_{k,l}}^T \mathbf{x})} \quad (15)$$

that is, the network consist of cn_f parallel linear filters whose outputs are the inputs to a soft-max nonlinearity; M is the size of the symbol alphabet and n_f s the number of filters per symbol. The network has as many outputs as symbols; obviously, in a binary case, $y_1 = 1 - y_0$.

It is not difficult to show that the softmax perceptron is equivalent to and RBF equalizer, in the sense that, for any RBF weights, there exists parameters $\mathbf{w}_{i,j}$ such

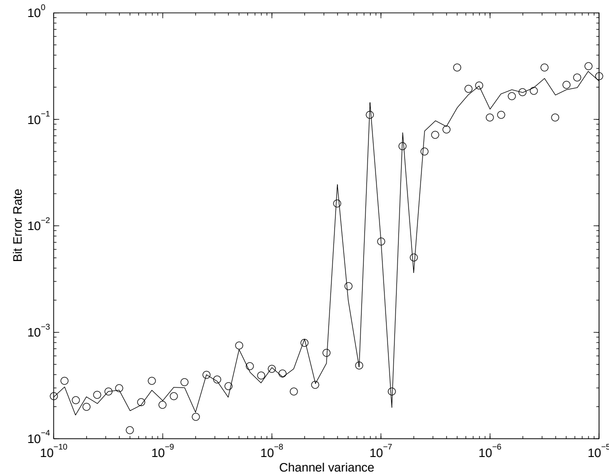

Figure 3: BER estimation in a time variant linear channel $H_1(z)$. Random walk model.

that both networks take the same decisions. However, the softmax outputs have a probabilistic interpretation (they lie between 0 and 1 summing 1 altogether), while the RBF networks is based on density estimation.

We explored the BER estimates based on softmax perceptrons for different cost functions given by Eq.(12) and (13). A similar performance was found for different values of α between -1 (cross entropy) and 0 (square error).

Figure 4 shows BER estimates based on the Softmax network with 4 filters per symbols for a binary transmission through linear channel $H_2(z)$ as a function of the SNR. An adaptation step μ_k decreasing with time according to

$$\mu_k = \frac{\mu_0}{1 + k/k_0} \tag{16}$$

was used, with, $\mu_0 = k_0 = 20$. The simulation shows

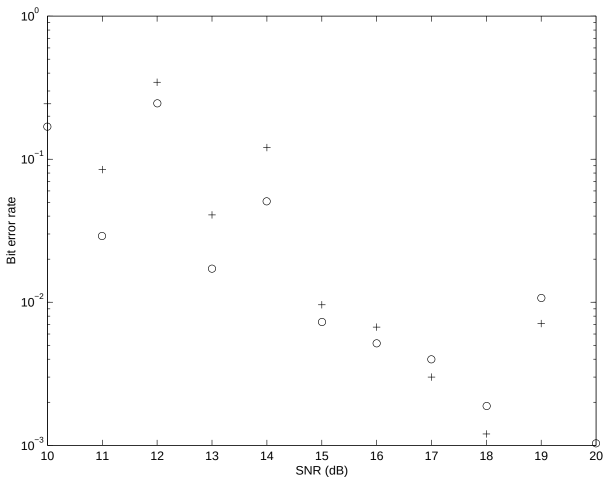

Figure 4: BER estimation of a Softmax Perceptron with 8 filters ("o"). Linear channel $H_2(z)$, 25000 training samples and 25000 test samples. The true error rate is indicated with "+".

that neural networks can be applied to BER estimate in the receiver; again, we found that the equalizer detects wrong training situations. However, RBF based networks requires less training samples: 25000 symbols have been used to train the equalizer of the example above; more complex channels and lower error rates require an excessive number of training samples. In order to find efficient BER estimates in time variant channels, faster algorithms may be required.

5 Conclusions

BER estimation in reception is essential to evaluate the communication performance in real time, and to detect wrong training cases in adaptive equalizers. We have shown that non-linear equalizers can be used to estimate the BER during data reception, without using any reference signal. We propose a simple estimation method that can be applied to any equalizer that computes posteriori probabilities of the transmitted symbols. Two kind of structures have been explored: density-based detectors, as the RBF and RRBF equalizers, and general feedforward neural nets trained with the appropriate learning algorithm. In the former case, it is found that BER estimates can be applied even in time variant channels, and Importance Sampling techniques can be used to reduce its variance. On the other hand, feedforward neural networks avoid making unnecessary assumptions about the channel response, although they require longer training sequences.

References

[1] S. Chen, G.J. Gibson, C.F.N. Cowan, P.M. Grant, Reconstruction of Binary Signals Using an Adaptive Radial Basis Function Equalizer; *Signal Processing*, Vol. 22, No. 2, pp. 77-93, 1991.

[2] J.C. Chen, D. Lu, J.S. Sadowsky, K. Yao, On Importance Sampling in Digital Communications (Parts I & II); *IEEE Transactions on Selected Areas in Communications*, Vol. 11, No. 3, pp. 289-308, Apr. 1994.

[3] J. Cid-Sueiro, A.R. Figueiras-Vidal, Recurrent Radial Basis Function Networks for Optimal Blind Equalization; *Proc. of the 1993 IEEE-SP Workshop on Neural Networks for Signal Processing*, pp.562-571; Baltimore, MA, June 1993.

[4] J. Cid-Sueiro, A.R. Figueiras-Vidal, Recurrent Radial Basis Function Networks for Optimal Symbol-by-Symbol Equalization; *Signal Processing*, Vol. 40, No. 2, Nov. 1994.

[5] J. Cid-Sueiro, A.R. Figueiras-Vidal, Cost Functions to Estimate Class Probabilities; *Procs of the Ist European Conference on Signal Analysis and Prediction*, pp. 113-116; Czech Rep., June 1997.

[6] J.F. Hayes, T.M. Cover, J.B. Riera, Optimal Sequence Detection and Optimal Symbol-by-Symbol Detection: Similar Algorithms; *IEEE Transactions on Communications*, Vol. COM-30, pp. 152-157, Jan. 1982.

[7] R.A. Iltis, J.J. Shynk, K. Giridhar, Bayesian Algorithms for Blind Equalization Using Parallel Adaptive Filtering, *IEEE Transactions on Communications*, Vol. 42, No. 2/3/4, Feb./Mar./Apr. 1994.

[8] M.C. Jeruchim, P. Balaban, K.S. Shanmugan, *Simulation of Communication Systems*, Plenum Press, New York, NJ, 1992.

[9] P.Smyth, J.W.Miller, R. Goodman, Objective Functions for Probability Estimation, *Procs. of the Int. Conf. on Neural Networks*, pp. 881-886, 1991.